多层次模型分析导论

■Ita kreft Jan De Leeuw 著 ■邱皓政 译 ■郭志刚 校

Introducing Multilevel Modeling



万卷方法 总 策 划：崔祝　雷少波
责任编辑：雷少波　罗杉
封面设计：余军

DUOCENGCI MOXING FENXI DAOLUN

本书是多层次模型分析的入门性读本，对于亟需了解多层次模型有效运用的研究人员或学生来说，本书是不可错过的。

多层次取向在社会、行为、教育研究的重要性日益提高，尤其从当代的发展趋势来看，此种模型比传统的回归模型提供了更贴近真实，且具启发性的分析与发现。

本书的作者不仅带领研究者去探讨实务上的议题，也对多层次模型执行中的理论问题进行了深入的讨论。同时，作者也利用真实的数据库，示范了多层次模型的研究设计与分析策略。

ISBN 978-7-5624-4060-4



定价：35.00元

多层次模型分析导论

■ Ita kreft Jan De Leeuw 著 ■ 邱皓政 译 ■ 郭志刚 校

Introducing Multilevel Modeling

重庆大学出版社

图书在版编目(CIP)数据

多层次模型分析导论/(美)克雷伏特(Kreft, I.),(美)里夫(Leeuw, J.)著;邱皓政译. 一重庆:重庆大学出版社,2007.4
(万卷方法)
书名原文:Introducing Multilevel Modeling
ISBN 978-7-5624-4060-4

Ⅰ.多… Ⅱ.①克…②里…③邱… Ⅲ.统计模型—高等学校—教材 Ⅳ.C8

中国版本图书馆 CIP 数据核字(2007)第 041897 号

多层次模型分析导论

Ita Kreft
Jan De Leeuw 著
邱皓政 译
郭志刚 校

责任编辑:雷少波 罗 杉 版式设计:雷少波
责任校对:邹 忌 责任印制:张 策

*

重庆大学出版社出版发行
出版人:张鸽盛
社址:重庆市沙坪坝正街 174 号重庆大学(A 区)内
邮编:400030
电话:(023)65102378 65105781
传真:(023)65103686 65105565
网址:http://www.cqup.com.cn
邮箱:fxk@cqup.com.cn(市场营销部)
全国新华书店经销
自贡新华印刷厂印刷

*

开本:787×1092 1/16 印张:11.75 字数:230千
2007 年 4 月第 1 版 2007 年 4 月第 1 次印刷
印数:1—3 000
ISBN 978-7-5624-4060-4 定价:35.00 元

原序

这一本书是写给社会科学领域中,没有很深的统计与线性代数基础,但是对于传统的线性模型,例如 **ANOVA** 与回归有一定认识的研究者或学生。

本书主要是以随机系数模型来处理多层次模型,这种模型具有固定变量与随机系数。第 1 章介绍了这个模型的基本概念与专有名词,并利用一些研究范例来说明其使用时机。第 4 章则以 **NELS88**(**National Education Longitudinal Study of** 1988; 详细的编码请见附录)的实际数据来进行分析与解释的示范。对于这个数据库,读者可以从下列网站得到(**ftp://ftp.stat.ucla.edu/pub/faculty/deleeuw/sagebook**)[1]。本书所使用的软件是由英国伦敦大学教育学院研究人员所发展的 **MLn**,可以处理多层次的数据。本书的实作部分,我们会以 **MLn** 的符号和语法来转换各种不同模型的方程式。在书中,**MLn** 的语法是放在方框当中,随着模型的变化,列出语法的变化,这些语法非常容易理解与阅读,读者并不需要花费太多力气去记忆指令和符号,我们会适时地加以说明。本书对于 **MLn** 的使用者来说非常的实用,但是本书的内容并不只限于 **MLn** 的使用者阅读。

本书与其他著作的不同在于,它是从研究者的角度出发,书中所引用的例子都是可能实际用来检验的模型,透过报表,可以看到一些重要的数据与信息。从实际操作的角度来看,本书特别着重于问题的澄清与方法学素养的提升。对于随机系数模型,我们也是采取保留与审慎的态度,尤其是传统模型与多层次模型的适用性与使用时机问题,我们认为其中有着一些两权消长、各有优劣的空间。本书所提到的阶层性嵌套数据(**hierarchially nested data**)的一些分析流程与替代做法,都是基于我们在教学与研究过程当中的经验,以及我们在多层次模型的网络社群当中所讨论的心得。如果读者有兴趣参与讨论社群,可以寄发电子邮件到 **mailbase@mailbase.ac.uk**,主旨为"**join multilevel your name**"。

最后,我们要感谢 **UCLA** 的 **Mahtash Esfandiary** 与荷兰莱顿大学 **University of Leiden** 的 **Rien van der Leeden** 对于本书部分章节的审阅,他们的意见提高了本书的可读性。

作者的提醒

由于本书的目的在于介绍统计方法,读者必须了解背后的一些限制,只有在一

[1] 该链接已不复提供数据文件,经与作者联系,新链接为http://gifi.stat.ucla.edu/sagebook,读者可自行下载数据库,转换成所熟悉的数据格式来进行演练。——译者注

译序

Kreft 与 **de Leeuw** 的这本书,是一本十分有趣且内容丰富的小书。今年初过春节的时候,闲来把它翻了一遍,发现他们能够在不到 140 页中,把多层次模型分析的概念、操作与解释,乃至于一些重要的议题,不拖泥带水地交代清楚,十分佩服,因此决意把它译成中文书,让国内研究者与学生可以很快地进入多层次模型的世界。

唯一遗憾的是,我自己以及诸位同事在分析多层次模型时,多使用美国 **SSI**(**Scientific Software International**)的 **HLM6** 软件,但是 **Kreft** 与 **de Leeuw** 是以英国伦敦大学发展的 **MLn** 为主,因此在阅读范例与数据时觉得格格不入。为了兼顾翻译的信达雅,以及个人的偏好,并考虑市场上 **HLM6** 的高度占有率以及未来的普及趋势,原文中有关 **MLn** 的介绍、运用与结果虽都完全保留下来,但是我另外以 **HLM6** 软件重新把作者所提出的绝大部分模型进行分析,分析的步骤与结果,列在第 2、3、4 章的最后,有部分模型并不是多层次模型,我则以 **SPSS**13 来处理。有兴趣的读者可以按图索骥,利用本书所附的数据库进行演练,将会事半功倍。此外,本书列举了很多网络链接与文献,如果想要深入了解阶层线性模式(**hierarchical linear modeling**)或其他多层次模型的读者,可以自行搜寻有关的信息,相信会有丰硕的收获。

本书的翻译并不困难,最大的收获是我个人从中的学习与体验。虽然我很早就处理过多层次数据分析的问题,也在课堂上教授这些高等统计的应用,但是总是点到为止,并没有机会好好深入了解这门学问。或许可以归咎于分析工具普及与便利性不高,但是最大的障碍是"隔行如隔山"的学科隔阂,如果不是身在量化方法与计量领域的有利位置,我还真没有机会一探多层次模型分析的究竟,尤其是要把这些东西写给别人来看时,要求彻底了解的压力就更明确了,有趣的是,翻译这本书,原本是想把这种新兴的技术介绍给更多的朋友,到最后受惠最大的却是我自己。

如同过去一样,这本书的完成,还是要感谢家人的支持与身边一些朋友的鼎力相助,像温福星老师与林碧芳老师费心的校阅修订,他们的协助让本书的可看性与正确性提高了不少。

译完这本书的最大心得是,我们真的是活在一个知识爆炸的时代。科技发展的脚步实在惊人,新知识、新科技、新想法不断提出,令人目不暇接,如果自己一直停留在自己舒适的空间,真是会有生于忧患死于安乐的遗憾。就像我们"台湾统计方法学学会"的好朋友们一起努力探讨结构方程模式(**structural equation model-**

ing,**SEM**)的正确运用的同时,阶层线性模式又在学术圈快速发展。很感动的是学会的伙伴们十分能够理解这种发展趋势,大家在 **SEM** 的探究之外,又“分案”开启了另一个有关 **HLM** 的探索空间。除了我以外,近期内还会有其他的 **HLM** 相关著作推出,例如东吴大学国贸系的温福星教授将会有一本 **HLM** 的专著出版。我们看好 **HLM** 的一个主要原因,在于 **HLM** 的方法论与技术取向在社会科学研究中占有相当重要的地位,甚至可以用“相见恨晚”来描述我们的心情。

为什么说相见恨晚,因为我们早就应该使用 **HLM** 技术来处理多层次的数据分析了。社会科学的量化研究,除了实验方法之外,多是以问卷、量表搜集众人的意见与经验,样本的取得很难做到随机,因此一群群的受测者,就可能因为具有组内的同质性,必须利用阶层化分析技术来处理组内相关(**ICC**)的问题。我实在很难想象,如果不用多层次模型分析就会得到扭曲的结果的话,那么我们过去几十年来所从事的各种组织、教育、社会、心理学等社会科学研究,究竟产生了什么知识与发现。这早已超越型一错误或型二错误的决策观,这实实在在是一个严肃的方法学问题,无法回避,只能面对。

相见恨晚不要紧,怕的是再次擦身而过,那就会是永远的遗憾了。

邱皓政
2006 年春谨识于
辅仁大学心理学系
心理计量实验室

原序

这一本书是写给社会科学领域中,没有很深的统计与线性代数基础,但是对于传统的线性模型,例如 **ANOVA** 与回归有一定认识的研究者或学生。

本书主要是以随机系数模型来处理多层次模型,这种模型具有固定变量与随机系数。第 1 章介绍了这个模型的基本概念与专有名词,并利用一些研究范例来说明其使用时机。第 4 章则以 **NELS88**(**National Education Longitudinal Study of** 1988;详细的编码请见附录)的实际数据来进行分析与解释的示范。对于这个数据库,读者可以从下列网站得到(**ftp://ftp.stat.ucla.edu/pub/faculty/deleeuw/sagebook**)[1]。本书所使用的软件是由英国伦敦大学教育学院研究人员所发展的 **MLn**,可以处理多层次的数据。本书的实作部分,我们会以 **MLn** 的符号和语法来转换各种不同模型的方程式。在书中,**MLn** 的语法是放在方框当中,随着模型的变化,列出语法的变化,这些语法非常容易理解与阅读,读者并不需要花费太多力气去记忆指令和符号,我们会适时地加以说明。本书对于 **MLn** 的使用者来说非常的实用,但是本书的内容并不只限于 **MLn** 的使用者阅读。

本书与其他著作的不同在于,它是从研究者的角度出发,书中所引用的例子都是可能实际用来检验的模型,透过报表,可以看到一些重要的数据与信息。从实际操作的角度来看,本书特别着重于问题的澄清与方法学素养的提升。对于随机系数模型,我们也是采取保留与审慎的态度,尤其是传统模型与多层次模型的适用性与使用时机问题,我们认为其中有着一些两权消长、各有优劣的空间。本书所提到的阶层性嵌套数据(**hierarchially nested data**)的一些分析流程与替代做法,都是基于我们在教学与研究过程当中的经验,以及我们在多层次模型的网络社群当中所讨论的心得。如果读者有兴趣参与讨论社群,可以寄发电子邮件到 **mailbase@mailbase.ac.uk**,主旨为"**join multilevel your name**"。

最后,我们要感谢 **UCLA** 的 **Mahtash Esfandiary** 与荷兰莱顿大学 **University of Leiden** 的 **Rien van der Leeden** 对于本书部分章节的审阅,他们的意见提高了本书的可读性。

作者的提醒

由于本书的目的在于介绍统计方法,读者必须了解背后的一些限制,只有在一

---

[1] 该链接已不复提供数据文件,经与作者联系,新链接为http://gifi.stat.ucla.edu/sagebook,读者可自行下载数据库,转换成所熟悉的数据格式来进行演练。——译者注

些前提条件被满足的情况下，本书的目的才可能达成。

第一个条件是，读者必须了解此一方法跟其他方法一样，所能够提供的答案跟数据的本质与搜集的方法有关，对于复杂的人类世界而言，统计方法不可能是完美的工具。

第二个条件是，读者必须了解多层次分析是另一种用来找到数据背后的结构，进而能瞥见真相的策略。就像其他的策略一样，有时候可能会不管用。

第三个条件是读者必须了解本书所介绍的多层次模型是基于某些假设成立的情况，如果这些假设不成立，或许还是可以进行多层次分析，但是标准误、显著性检验等一些统计程序都可能会有问题。当然诸如斜率为随机的这种假设，可以通过样本数据来进行确认，但是如果研究结论是斜率为非随机，并不意味着这个假设一定是错误的，很可能只是在某一个样本上得到这种结果。可能真相是支持这个假设的，只是统计方法无法证明它而已。

第四个条件是，多层次模型的使用者必须了解统计模型是一种数学的模型，如果统计模型所建构出来的数据近似于作者所假设的状况，那么研究结论符合真相的机会会较大。事实上，由于研究者所探讨的问题的复杂性，统计模型的提出往往偏离真相。复杂的模型虽然可以模拟真相，但是由于过于复杂，却削弱了统计方法的效用。如何整理、分析复杂的数据不是一蹴而就的工作，复杂的统计模型尤其难以解释，而且不容易重复观测到相同的结论。复杂的模型先天上就会对于微小的变化非常敏感，造成参数估计因为小小的变动而有很大的不稳定性。

第五个也是最后一个条件是，多层次模型的使用者必须了解这种统计工具之所以有效，是因为数据具有多层次的结构，或是因为理论背景支持使用这种分析，或是因为我们对于数据本身有相当的了解。因此，在进行多层次分析之前，有必要对于数据进行先期的探索检验。

多层次模型这种复杂的统计模型或许是符合真实现象的模型，但是在数据探索的阶段我们并不建议使用，我们也不建议使用过于复杂的模型，例如带有许多解释变量的模型，或是带有复杂的跨层级交互作用的模型。

本书中，都是以小模型来进行示范，变量的选择都具有理论基础，对于数据的性质也有充分的掌握。

目 录

概　说　1

1.1　绪　论

本章的目的在于简介多层次模型的发展历史、使用时机,以及软件的应用。有关应用的部分,本章将从几个不同领域的研究实例来说明。针对阶层性嵌套数据分析的一些概念与名词也将在本章介绍。最后,则对多层次模型的历史进行简要回顾,并以几种软件的介绍作为结语。

1.1.1　阶层、宏观层次与微观层次

阶层数据结构(hierarchical data structure)在社会与行为科学领域中非常常见。如果个体身处不同的组(团体),此时有些变量与个体有关,有些变量则与团体有关。例如在学校所搜集的学生资料,可能包含一些用来描述学生特征的变量,如社经地位、对于写作业的态度、性别、种族背景;还有一些变量则在描述学校,例如公私立之别或学校类型。学校效能(school effectiveness)的研究者在进行他们的研究时,同时会搜集这个层次的数据,藉以了解学生微观层次与学校宏观层次对于学习成绩的影响。这种在学校所从事的研究的例子以及其他许多类似的状况,说明了我们需要一种可以同时处理多层次测量数据的各层次数据的技术。

多层次模型被设计用来分析阶层结构的数据。在进入详细的介绍之前,我们先将"阶层"一词加以介绍。所谓阶层(hierarchy)系由较低层次观察数据嵌套(nested)[1]在较高层次之内的这种数据结构所组成。例如学生嵌套在学校之内、员工嵌套在公司之内、重复测量嵌套在个体之内。最低层次的测量称为微观层次(micro level),其他高层次的测量则属宏观层次(macro level)。宏观层次通常系由不同的组别(groups)构成,更正式的说法是不同的脉络(contexts)。因此所谓脉络

1　nested 除了被译成嵌套以外,亦可被翻译成内属、嵌套、内嵌等。在本书当中,这几个翻译可互换使用。——译者注

模型(contextual models)一词是指兼具微观与宏观层次数据的模型。脉络模型有时仅有两个层次,例如学生(微观层次)嵌套于学校(宏观层次)之内,亦有可能超过两个层次,例如学生嵌套于班级,而班级则嵌套在学校之内。当然研究者可以想到更多的层级,例如学生嵌套于班级、学校、州别、国家地区等。一旦能够理解阶层的特性,我们可以发现它几乎无所不在。

1.1.2 多层次模型

如果一个模型包含了不同层次的测量变量,称为多层次模型(multilevel model)。在多层次模型中,各脉络可估计出一条低层次的个别直线方程式。在这条回归线中,通常各脉络都有相同的解释变量(explanatory variables)[2]与结果(outcome)[3],但是有不同的回归系数。这些个别回归方程式被一个高层次模型所联结,在高层次模型中,第一层次的回归系数可被第二层次的解释变量所解释。

用来联结这些个别回归方程式的高阶模型的特性,决定了整个模型对于数据分析的方式。在实务上有多种处理方式,最开始的起点是没有高层模式来整合个别方程式的模式,此时每一个脉络各拥有一个回归方程式。虽然这是很自然应该为之的做法,但是从统计的观点来看,此一做法并没有任何新意。

进一步讲,将第一阶层各脉络的回归系数作为第二阶层的反应变量(response variable),称为"斜率结果(slope-as-outcome)"分析(Burstein,et al. ,1978)。从统计的观点来看,在各组内与各组间所进行的回归分析彼此并无关联,他们仍是个别进行的分析。不论是未经联结或已被高层模式联结的分析,回归系数都是固定数值(fixed),而非随机变动(random)。如果一个模型利用了全部数据进行分析,则被称为变动系数模型(varying coefficient model),此种模型的分析方法即如其名,是把每一个组分开进行估计,因此每一组拥有各自的一组回归系数。

每一个组进行个别回归分析之后,再以高层解释变量来解释第一层系数的这种说法,尚不足以说明多层次模型的内涵。基本上,多层次模型的基本特性是将研究者所关心的特定层次的数据,用不同的模型通过统计的整合来加以分析。

最简单的整合模型是随机系数模型(random coefficient model)。这种模型中,第一层的回归系数在第二层被以随机变量来处理,这意味着第一层回归系数是从某个机率分配取样而得,此一分配最重要的参数——平均数与方差,可从多层次模

[2] 解释变量即为自变量(independent variable),在回归分析中又称为预测变量(predictor or predictive variable)。——译者注

[3] 结果即为因变量(dependent variable),在本书多以反应变量(response variable)称呼之,有时则称为结果变量(outcome variable)。——译者注

型中估计得出的参数所获得。一般来说，在随机系数模型中加入第二阶层的解释变量是一般通则性的做法，也是非常有用的策略，因此一般被通称为多层次模型。

在本书中，将组(groups)与脉络(contexts)视为同义词而交互使用，是指其为阶层数据当中第二层(或较高层)的分析单位。为避免混淆，本书所使用的组别概念有别于实验心理学所惯用的实验组与对照组的组别概念。本书所谈的组别是自然形成的归类结果，例如学校或公司。而脉络则是指与组别相同的概念，而非涉及社会学所谈的更广义的概念。在以下的章节中，我们将以实际的例子来说明阶层性嵌套数据的形貌。

1.2 范 例

在本节当中，我们将介绍一些在不同领域涉及阶层嵌套数据的实际范例。第一个例子是企业员工嵌套在公司之内。研究发现指出，在不同阶层所进行的不同层次分析所得到的结果不尽相同。在这个例子中，我们会讨论到各脉络的观察值具有相依性。第二个例子则是学生嵌套于学校的班级之内。其他的例子则为不同领域的实际范例。

1.2.1 企业员工的薪资水平

第一个例子是 Kreft 等人(1995)所进行的研究。研究资料是从 12 个不同的公司所获得，微观层次的解释变量是教育水平，反应变量为薪资收入。公司的类型，例如公营与民营，为第二个阶层解释变量。对于员工层次的分析发现，教育水平与收入具有正向关系，教育水平较高者，薪资收入较高。以 12 个公司为观察对象的公司层次分析发现相反的结果：教育水平与薪资收入呈现负向关系，当一个公司的平均教育水平越高时，该公司员工的平均薪资收入越低。

在此例中，公司层次变量为聚合测量(aggregated measurement)，也就是以公司的平均教育水平为解释变量，反应变量是公司的平均薪资。分析结果显示，不同层次的分析得到不同的结果。高层次聚合数据的分析结果与原先微观层次的结果不同的这种现象，在 Robinson(1950)的研究中即已发现，此种聚合偏误现象称为 Robinson 效果(Robinson effect)。前面的组织研究数据中，微观层次的分析发现教育水平对于收入有正向效果，但宏观层次的分析发现负向效果，逻辑上来说，两个层次的教育水平变量是在测量不同的东西，取决于分析的单位。此一结果显示我们有必要将两个层次的分析同时加以处理，因为两个层次的结果都非常重要，而且彼此具有关联。

此一范例显示出阶层嵌套数据的一个重要特性,亦即同一个公司的员工比起其他公司的员工更为相像。公司内员工的同质性可以通过公司的特征来反应,更具体来说可由组内相关(intra-class correlation)(注1)来衡量。如果组内相关很高,团体内具有同质性,而且可能与其他团体的差异很大。此一现象也出现在本范例的结果中,亦即教育水平对于员工收入的解释,在公营企业远高于私人企业。一般而言,如果组内相关较低时,团体间仅有些微差异。如果组内相关低至近乎0,则研究者所关心的变量无组间差异,换言之,同一组内的个体间差异,就像该组个体与其他组个体之间具有相同的差异程度。一个数值为0的组内相关意味着数据的丛集不影响研究变量关系,因此可以忽略组间的差异。如果组内相关存在,数据的嵌套结构就必须加以考虑。忽略组内相关会导致结果的信度问题,但是组内相关必须具有统计的显著性,并且有相当的强度,有关此点可参见 Cochran(1977)的详细讨论。

1.2.2 药物滥用预防研究

第二个例子为药物滥用防治研究(Kreft,1994)。在该药物防治研究中,研究者所关心的是高中青少年药物防治计划是否有效。实验处理是防治计划,测量与分析的对象是学生。变量则依不同的阶层,在学校/班级层次与学生层次中有所不同。参与研究的学校是随机抽样得来的,学生所属的班级也可以被视为是从某一类特性的学校所随机选择的样本。研究所测量的变量包括学生的风险因子,例如心理因素、学业成就、贫富差异等,在学校或班级层次的风险因子则包括一个学校药物滥用的程度、邻近区域药物滥用的状况。

有关药物防治的文献指出,个体风险因子与药物防治计划之间具有交互作用,而学校风险因子也与防治计划具有交互作用。文献中还可以看到许多有关脉络与学生特性的假设效果。若以多层次分析术语来说,这些关系应称之为跨层级交互作用(cross-level interaction),因为变量的关系横跨学校与学生层次。可以预期的是,某些学生(例如高危险群学生)容易受到某种环境的刺激而使用药物,但是在其他的环境下则可能减少他们使用药物的机会。为了检验这些研究假设,我们所需要的不仅是嵌套的数据结构,而且要能够估计跨层级的交互作用。

1.2.3 学校效能研究

第三个例子是学校中老师的教学效能研究,在这个例子中,阶层结构扮演重要的角色。研究所关心的对象是学校与老师,同时也包含了学生。研究者可能对于学生个人以外的因素,例如组织结构如何影响学生的学习成果感到兴趣,研究者也可能关心老师的教学经验、智能、教学风格等因素是否影响学生的学习。典型的例子是 Cronbach 与 Wenn(1975)、Burstein 等人(1978)与 Aitkin 和 Longford(1986)所

做的研究。

如果以实验方法来比喻，不同的老师或学校可以视为一种实验处理，此时协方差分析可以说是最常用的一种分析技术。但是应用 ANCOVA 在这个例子上可能出现一些问题，也是这个研究领域最重要的一个问题，亦即虽然研究者可以把脉络差异放入模型加以考虑，但是与脉络有关的特殊变量无法被放入模型中来进行解释。ANCOVA 所能回答的问题，是学校间是否具有差异，然而在学校效能研究中，这并不是最重要的问题。重要的是为何这些学校会有不同？这已超过 ANCOVA 所能够应付的情况。

学校效能研究所问的问题，通常是何种老师的特性或学校的组织特性会影响学生的学习？对于某些学生，是否小型的学校会比大型学校更能提升学生能力？是否私立学校会比公立学校对全体学生较有好处？那么私立学校为什么会比较好？学校的规模对于特定类型的学生，例如高危险群、不同性别的学生的影响如何？如果一个研究所关心的是特定的环境（学校规模）对于特定学生（高风险学生、不同性别学生）的影响，则必须使用特殊的分析方法来回答。基于实验方法所发展出来的 ANCOVA，最主要的一个问题是无法处理组内相关问题。在实验设计中，受试者被随机分配到不同组，各观察值被假设是相互独立。但是在真实世界的观察数据中，同一组的个体具有共同的经验，导致相同脉络的个体之间具有相依性，这个问题必须加以处理。

1.2.4 临床治疗研究

另一个与多层次分析有关的领域是临床心理学，特别是对团体治疗研究的评估研究。在团体治疗中，治疗型态是由研究者控制的影响因子，但是团体动力则不是研究者所能够控制的变因。对于个案以团体方式接受治疗，此时每一个团体中的个案分配可以说是一种随机选择的程序，亦即接受团体治疗的个案一开始可以被视为是相互独立的个体。但是随着时间的经过，团体的状态可能会有所改变。每一个团体内部的互动受到团体动力的影响，而团体动力则随着时间产生无法预期的变化。

如果现在有指导式团体与非指导式团体两种不同的治疗方法，使用相同方式的治疗团体可能会发展出不同的团体动力型态，尤其是非指导式的团体所发展出来的团体动力更无法预期。基本上，每一个接受治疗的个案的行为，一方面反应了治疗的效果，同时也受到该治疗团体所发展出来的团体动力的影响。团体成员的互动将使每一个团体内的成员相似性高于其他团体的成员，因而使得团体成员之间不再有统计上的独立。

如果我们使用传统的方法来分析这种数据，显而易见地会出现问题。传统的

方法是固定效果的 ANCOVA。前测是接受治疗个案的心理功能,作为协变量,后测分数则作为反应变量。各治疗团体嵌套于治疗方法之中(不是指导式,就是非指导式)。以 ANCOVA 来分析,会忽略了随着时间逐渐发展出来的团体动力影响(组内相关),导致参数估计值的误差方差被低估。各团体虽然嵌套于治疗方法,而且一开始各团体是相同的状态,但是随着时间的变化,发展出实验者无法控制的团体动力,这是无法被传统的 ANCOVA 模型所处理之处,也无法被视为治疗的效果来解释,因此需要一种新的方法来处理这种组内观察值的相依性。而各组在模型上的差异,通过宏观层次的不同团体特征(例如治疗者的特性或团体动力)的影响,也必须由特殊的分析方法来处理。第 1.2.8 节我们将讨论其中一种特殊状况。

1.2.5 成长曲线分析

在阶层数据分析中,分组并不一定是一群个体的集合的这种自然社会脉络现象。个体也可能是宏观层次的观察单位,而不必然是微观层次的数据。如果今天所分析的数据是个体重复观测的数据,研究者所测量的数据是嵌套在个体之内,而且彼此之间具有相关。此时,组内相关所测量的是同一个个体后面的行为与他自己先前的行为之间的相似程度。有关重复测量随机系数模型分析的书可参考 Lindsey(1993)与 Diggle 等人(1994),但是在第 1.5 节所介绍的几本讨论多层次模型基本原理的书,多半都有专章介绍重复量数的数据分析。

在教育研究的例子中,如果对学校的学生具有重复量数的测量,即形成了一个三阶层的数据结构。重复量数是第一层,嵌套于学生之中,成为第二层,而学生则嵌套于学校之中,成为第三层。

在第 5 章中,我们可以看到多层次的技术可以非常容易地处理重复量数的成长曲线数据,包括了带有遗漏值的不平衡性数据。虽然在社会科学领域多半重视的是横贯数据,但成长曲线分析(growth curve analysis)可以说是多层次技术最特别的一种应用。

1.2.6 地理信息系统

有关空间统计学的应用越来越多,例如公民投票、选举资料、调查研究通常会以地理因素来加以组织,但是直到最近,学术领域才发展出适当的分析工具来处理地理的信息,而多层次技术即是其中一种可以用来处理地理信息的技术。例如城镇或个人可以嵌套于地理区域中,因此组内相关可以用来反应同一地理区域内的空间自相关(autocorrelation)。Cressie(1991)的著作非常详尽清楚地整理了空间统

计的分析方法,其中也包括了阶层线性模型与克利金法(kriging)[4]之间的关系。

同样的,多层次分析技术可以很容易地应用于地理信息系统(geographical information systems)的空间分析工作。在此,我们引用McMillan与Berliner(1994)的研究,该研究测量了爱荷华州的3 842个农场的玉米产量,这些农场位于88个郡,依其大小区分为三种不同规模。McMillan与Berliner(1994)的研究基本上是方差成分模型,农场是随机成分,88个郡也是随机成分。各郡的协方差矩阵可以利用马尔可夫随机场技术(Markov random field techniques)来处理。本书所介绍的其他形式的多层次技术也可以作为替代模型。也就是说,88个郡的特征作为第二层解释变量,来解释88条以农场为分析单位的回归方程式的回归系数。不同郡的相邻农场并没有空间相关,同一个郡的农场则有空间相关。

1.2.7 元分析

在元分析(meta-abalysis)[5]中,不同研究的结果被整合在一起再加以研究,每一个研究会产生一个效果量或差异量的估计值,可被拿来分析。在元分析中,通常个别观察者的数据不会再被拿来分析,个别数据不是不可考了就是被忽略了。元分析的数据型态是百分比、平均数或相关系数的资料。对于这些数据,我们仍然可以利用多层次分析技术,因为研究者所搜集的各个独立的研究或文献,可以被适当地归类。虽然个别的数据已不存在,但是文中仍可找到估计标准误的讯息。这些标准误的数据就宛如是多层次分析的第一层估计后的结果。第二层次的分析就是研究与文献的元分析的部分。详细的细节可以参考Bryk与Raudenbush(1992)的著作。

1.2.8 双生子与家庭研究

这个例子与前面的各范例不同,在于组的规模非常小。在双生子研究(twin study)中,我们可能搜集了许多的样本数为2的双胞胎数据,这么多组的双胞胎,每一组的规模都很小。在各组之内,由于只有两个观察值,因此无法估计各自的模型,此时统计的稳定性来自于数目较大的组数。

在心理与教育领域,长期以来有着IQ受先天或后天影响的辩论,此一争议有关于IQ为先天遗传决定还是后天环境影响的结果。若我们搜集到同卵双生与异卵双生的双胞胎数据,如果IQ是基因遗传所决定,同卵双生的组内相关应该很高,

[4] 克利金法是一种由地理统计学为基础所发展来推估未知参数的内差方法,为最小均方根差方法的应用之一,可以利用不规则的数据值内插成规则网格点上的数据值,在实际应用上常作为绘制空间等值曲线(等高线、等水位线)的方式。——译者注

[5] 元分析(meta-analysis)又可译为后设分析。——译者注

而与异卵双生的组内相关应有所差异。但若以一般的理解来看,IQ 的遗传影响力应该是相当的。由此类研究可知,虽然有时我们的数据库的组内样本规模很小,但是有很高的组内相关。在多层次分析中,我们可以利用不同的解释变量来分别解释微观层次的观察值与在宏观层次上的配对数据。

在社会与行为科学中,另一种类似的研究是配偶数据的研究。在有关配偶的研究中,每一组的观察单位也是两个个体,形成一对夫妻。所不同的是在双胞胎研究中,不分哪一个个体为 A 或 B,两个人可以互换,但是在配偶的研究中,一个为夫,另一个为妻,组内数据的对称性低于双生子研究,但是却是符合社会真实状况的数据型态。

1.3 综述与定义

在前面的章节中,已经介绍了一些阶层结构数据分析的重要名词与概念。为了便于后续章节的说明,以下我们将针对脉络模型、组内相关、随机与固定系数、缩动、预测和跨层级交互作用几个重要名词加以定义说明。

1.3.1 脉络模型

传统上,脉络模型被定义成包含有两类变量的模型:微观层次变量与聚合脉络变量(aggregated context variables),例如团体平均数或中位数。当我们搜集到学生的数据,而学生嵌套在学校之内,学生的学习成绩可以被社经地位(SES)所预测,同时也被各学生所属的学校的社经地位平均数所预测。此时 SES 变量被使用了两次,第一次是个别学生层次变量,另一次是整合后的学校特征。在文献中,回归分析若包含此种聚合脉络变量则称为脉络模型(Duncan et al. ,1966)。

本书对于脉络模型的概念,扩大了传统的观点。脉络模型是指在任何一种线性回归模型中,若包含了低层次变量与高层次变量,高层次变量则可能是整合后的脉络变量,也可能是其他类型的整体特征的测量结果。整体特征(global characteristics)被定义成对于脉络特性直接加以测量所得的变量(参见 Lazarsfeld & Menzel, 1969),而不是从微观层次测量得到的变量所整合得出。任何具有个体变量与脉络层次特征的回归模型,在本书中均视为脉络模型。基于此一定义,带有高层次变量的随机系数模型也是一种脉络模型。

1.3.2 组内相关

有关组内相关(intra-class correlation; ICC)的概念,我们以学校效能研究的范例来说明。当学生嵌套在学校之内,学生与学校都可能是研究者所关心的对象。

基本上，在同一个时间与空间所搜集的观测数据会较相似，在不同时间与空间所搜集的观测数据则较有差异，因此，基于生活经验与环境的相似性，同一个学校的学生会较相似，不同学校的学生的状况则较不同。相同脉络的共同性导致了观察数据的相依性。

组内相关即是组内的个体间相依程度的测量。当个体间因为时间与空间因素导致有较高的相同经验与共通性时，彼此的相似性提高，当相似性高到了某种程度时，每一个人甚至可以说是完全相似的复制品。最高的相似性通常发生在当两个个体是双胞胎，或是在同一个家庭中成长的小孩。另一个常见的高度组内相关的例子是同一个个体的重复测量。

检验资料中是否具有高度的组内相关是一件重要的工作，因为组内相关会影响传统线性回归模型的误差方差。误差方差代表了所有遗漏的解释变量与测量误差的影响，而且这些误差具有相互独立之基本假设。在传统的线性模型中，解释变量的遗漏被假设为一种随机、非结构性的现象。但是当数据具有丛集结构的特性时，此一假设是受到威胁的。例如，在学校效能的研究中，学校气氛或同侪压力可能是一个没有被测量到但具有结构性影响的变量。同一个学校或班级的学生的误差项协方差的幅度，即会以组内相关系数的型态浮现。

组内相关一般以 ρ 表示，可以以几种方式来解释。前面是以相同的时空下，个体所共同经验的程度来界定，此时可以被称为团体同质性(homogeneity)指标。组内相关更正式的定义，是指当数据具有两个层级的结构时，高层次观察单位之间的方差(组间方差)占结果变量方差的比例。不论从哪一个方法来界定，组内相关所反应的都是组内(或脉络内)相依性的存在，这意味着如果组内相关存在时，表示这一群个体是属于一个丛集，此时，传统线性模型对于观察值需为独立的假设即遭到违反。

在文献上，当组内相关存在所造成违反上述假设的影响，会反应在型一错误率(alpha 水平)的增加。当组内相关存在时，30 个在同一个班级的学生已非 30 个独立的个体。学生之间独立性的流失，取决于班级内个体的相似性或是组内的同质性的高低。组内相关的强度决定了观察值间真实存在的独立性。由于显著性检验是以观察值的独立性为基础，当组内相关存在时，将造成传统线性模型的显著性检验过于宽松(参见 Barcikowski, 1988)。Barcikowski 指出了在多数的 ANOVA 检定中，参数估计的标准误都是被低估的，一个很小的组内相关(假设 $\rho = 0.01$)，即可能造成型一错误率的实质扩大，如表 1.1 所示(表 1.1 是以 ANOVA 的情况来说明)。对于一个大样本($N = 100$)，$\rho = 0.01$ 的低度组内相关即会使型一错误率由 0.05 提高到 0.17；对于一个小样本($N = 10$)，$\rho = 0.20$ 的组内相关会使观察到的型一错误率提高到 0.28，而非原本假定的 0.05。一般来说，当一个大样本具有低度

的组内相关,与一个小样本具有高度的组内相关,对于型一错误率的膨胀具有相同的影响力。然而,低度组内相关对于小样本的型一错误率的影响虽然较小,但是随着样本数增大,对于型一错误率膨胀的负面影响程度便会扩大。

表 1.1　当组内相关存在时,$\alpha=0.05$ 的型一错误率的膨胀现象

组样本数 Ni	组内相关 ρ		
	0.01	0.05	0.20
10	0.06	0.11	0.28
25	0.08	0.19	0.46
50	0.11	0.30	0.59
100	0.17	0.43	0.70

表中的数值为观察到的 alpha 水平;Ni 为组内的观察值数目。

组内相关对于型一错误率的影响程度,类似于传统相关系数的特性:低度的相关系数在小样本中不显著,但是在大样本中会显著。有关这些议题的讨论,可以参考 Barcikowski(1981)与 Cochran(1977)的著作。

1.3.3　固定与随机系数

一般人对于“随机(random)”与“固定(fixed)”两个名词有诸多混淆之处。在线性模型当中,这两个名词有三种不同的说法:随机与固定效应、随机与固定变量、随机与固定系数。

固定与随机效果的概念普遍应用在实验研究当中,涉及实验处理与实验组的设定,这些数据多以方差分析(ANOVA)来分析。A 因子若为不同的实验处理,当研究者所关心的每一种可能的实验状况都被包含在 A 因子中时,A 因子被称为固定效果(fixed effect)。随机效果(random effect)则是指当 A 因子中的各个实验处理是所有可能状况的一个样本。有关固定与随机效果的经典论述,可参考 Scheffe(1956)与 Wilk & Kempthorne (1955)的文章。

例如,今天有一个精密控制的实验室实验研究,用来检验某一种药物的药效。实验组是施以该药,控制组则是给予安慰剂。实验过程中,受试者的一些相关反应被搜集作为反应变量。很清楚,这两个实验组形成了一个固定效果,因为这个研究目的的本质仅涉及这两种状况。但是在真实环境中,以真实生活当中的分组状态所进行的准实验研究,固定实验处理的假设几乎无法完全符合。前面所举出的学校效能研究的例子即说明了这个难题。不同型态的药物滥用防治计划的实验研究所使用的受试者为真实的学校学生也无法符合实验室实验随机化的要求,各组的

受试者并非完全相等,参与研究的各校实际上是从所有可能参加研究的学校当中的一个随机样本。药物滥用防治计划对于某个学校的影响理应被视为随机而非固定。因此实验效果应为随机效果而非固定效果。

固定效果与随机效果的区分是非常重要的,因为所造成的推论结果和研究发现所能够推论的范围有别。例如,固定效果仅允许对于研究当中所发生的各种状况的推论,"效果"的本身是被假设为固定常数而无测量误差。相对的,在随机效果模型中,例如前面所提及的学校效能研究,推论的范围可以扩及到参与实验的学校以外。此时,研究者的目的是对学校母体的推论,而非实验当中的学校。实验的效果并不假设为常数,而是具有些微差异,或带有抽样误差的测量。这是因为被选择的学校是抽样得来所导致的结果,我们所欲推论的是母体的本身。我们可以预期若以另一组样本重复研究,得到的结果多多少少会有差异。

相同的,常数与随机两个名词在对一个变量的描述上也是有预设的定义方法,背后同样也涉及了测量抽样误差的问题。对于一个变量来说,随机化一词与随机系数模型当中的"随机"是不同的概念。随机系数模型假设变量是固定的。我们在这里讨论随机变量一词,目的在于区分随机变量与随机系数。

固定与随机变量是统计理论上的概念。对于随机变量(random variable)较松散的定义,是指一个变量的数值是来自于机率分配。因此,一个随机变量会有一个期望值(亦即平均数)与方差(可能是已知或未知)。一般来说,我们会假设随机变量在测量时是带有抽样误差,每一次测量都不一样,IQ 的分数就是一个例子。一个人的智力可以在同一个情况下反复多次测量,每一次都得出不同的数值。相对的,固定变量(fixed variable)表示一个变量的数值是已知的,是固定的数值。例如性别,每一次某一个人被测量他的性别时,我们假设相同的性别将会出现。传统上的回归模型中的预测变量与方差分析当中的设计变量(design variables)通常都被视为固定变量,表示该分析具有固定的设计。但是在线性结构关系的分析中,变量多都被视为是随机的。以松散的定义来说,固定变量所关心的,是变量当中所出现的那几个数值,随机变量所关心的则是参数背后所存在的机率分配。在本书中,解释变量皆假定为固定变量。

至于固定系数(fixed coefficient)与随机系数(random coefficient)和线性模型的参数特性有关,是伴随着随机系数模型而来的新名词。在传统的回归模型中,回归线当中所估计的参数包括斜率与截距。传统上,这些系数均被假设为固定数值而不会变动,系数数值是从观察数据中所估计得出。而随机系数是指系数的数值服从某一概率函数分布,一个机率分配即有其期望值与方差。在多层次模型架构中,第一阶层的回归模型系数被用随机系数来处理。有时研究者的兴趣在于这些参数的期望值,有时则是这些随机系数的方差,有时是两者。

以斜率来说，斜率的随机系数被区分为两个部分：第一是整体斜率（overall slope）的数值，是从所有的个体中所估计得出，不在乎这些个体属于哪一个组。第二是斜率方差（slope variance），表示每一个组的斜率与整体斜率有所差异的变动情形。在多层次模型中，若为随机系数模型，则允许各组从平均数处存在变异，不论是截距或斜率上的变异，进而可以估计变异情形。在本书中，我们将对这些模型的各种变化加以详细说明。

1.3.4 跨层级交互作用

跨层级交互作用（cross-level interaction）是指在阶层结构数据中，不同层级变量对于反应变量的交互作用。例如学生与脉络（学校）之间的交互作用，学生的特性如性别，脉络的特性如老师对于性别的态度。在教育领域，Cronbach 与 Webb（1975）最早提到了跨层级的交互作用。

Cronbach 与 Webb（1975）的研究假设认为，一个有效能的教师并不是针对所有的学生有效，而是针对某一种学生。如果某一种老师对于资优学生有较佳的效能，那么意味着这些学生个人的性向与学习成绩表现之间的关系受到这种老师的强化影响。我们可以说这一种老师拥有精英领导式（meritocratic）教学风格。另一方面，如果某种教师较为擅长指导学习迟缓的学生，那么性向与学习成绩之间的关系就可能被弱化，我们可以说这一种老师是一种平等主义式（egalitarian）的教学风格者。前面一种教师可能会强化高低成就学生之间的差距，后面一种教师则可能缩小这种差距。在教育的文献上，将这种效果称为性向与处理交互作用（aptitude-treatment interaction; ATI）。此时学生属于微观层次，老师是宏观层次，老师与学生的交互作用即微观与宏观的跨层级交互作用。

1.3.5 预 测

对于一个回归模型，不论是固定系数或随机系数，都具有预测（prediction）与解释（description）的功能。预测可能是实际现象的预测，或是虚拟状况的预测。如果是实际现象的预测，研究者可以从实际发生的状况来检验预测效果的好坏。如果是虚拟的预测，研究者所假设的状况无法通过实证数据来检验。例如，研究者预测，在其他变量维持不变的情况下，父母的社经地位每增加一个单位，学生在 SAT 考试将会增加 40 分。很明显，此种研究者所构思出来的实验研究无法在真实世界中验证，也无法从数据的操作来验证假设预测的对错。

回归模型也可以用来描述变项间的关系。我们可以计算出在特定社经地位水平的学生的 SAT 平均成绩，此分数与社经地位分数之间的关系为线性函数。更精确地说，这些组内学生分数的变异归因于回归（due to regression）。再一次的，如同“预测”一样，我们可以看出这些“描述”基本上仍有其限制，无法通过实验程序来

验证。

统计的角色是在取代重复实验的角色。背后的基本假设，是这些实验若被重复一百次，每一次的结果会有些许的不同。更精确地说，在所有的100次观察值中，有95%不会超出真值的两个标准误以外。值得注意的是，这些理论上的"重复"仍是一种无法被实现的实验。

在这个脉络之下，多层次分析的目的是去得到更有效的预测，同时也在正确描述数据所呈现的关系。我们必须了解，多层次分析通常可以通过引入额外的参数或方差与协方差(第一层次观察单位间的关系)成分，来改善对于变项关系的描述。然而，增加额外的参数必须付出代价，因为回归系数的估计会失去原先的精确性。事实上，精确性(precision)一词在我们追求估计的无偏性时已经不是最重要的考虑。在统计的世界中，处处可见到此种两权消长(trade-off)的现象。在这里所遇到的消长问题是增加参数但各参数估计的精确性会有所减损。如果在一个模型中，有太多的参数需要估计，研究结果的精确性将会大幅度地降低，造成预测上的无效。

1.3.6 缩动与借力

随机系数模型是两种模型的折衷妥协版：其一是每一个脉络分别建立自己的模型，另一种状况是所有的脉络同时使用同一个模型。前面的一种状况是非常浪费奢侈的做法，因为每一个脉络都要进行参数的估计，每一个模型至少有三个参数：截距、斜率、误差方差。进一步的，不同的脉络需要不同的分析这个做法与我们认为各组是有关联的认知是相冲突的，因为每一个学校都是身处于同一个教育体系。但以单一模型应用在每一个脉络是太过于简略的做法，因为没有足够的参数可用于描述我们的观察对象。

此一两权消长的现象在统计上十分常见。我们可以选择介于参数很多与参数很少之间的模型(参数的合理数目需与样本数进行比较)。如果我们所估计的参数数目较少，所估计的参数会有较高的精确度，但是可能付出得到有偏结果的代价。之所以会比较精确，是因为我们有较小的抽样方差；之所以会存在偏差(biased)，因为方差是围绕着不正确的估计值。当有较多的参数，我们可能会有较小的偏差，但是有较差的精确性，也就是有较大的标准误围绕着较正确的数值。过于极端都不足取，我们必须去找到一个最适切的居中平衡点。在很多地方都可以看到这类的讨论，例如 de Leeuw(1994)。

最近在国家研究委员会(National Research Council,1992)出现了另一种观点，认为每一个研究都可能因为规模太小，导致研究者对于所关切的现象无法得到足够精确的研究结果。但是，通过合并不同的研究或学校，我们得以从其他研究或学

校“借力(borrow strength)”,从而获得较为强而有力的结果。这种想法或许让某些人觉得惊讶,但这种作法也仅适用于当学校是研究者所关心的对象时。

“借力”的效果并不会改变单独的一个学校的个别估计值会有所缩动(skrinkage)这个事实。当个别脉络的观察值数目很少时,缩动的结果特别容易被看到。随机系数模型所计算得出的参数估计,是以全部脉络整合得出的估计值,以及以个别脉络分别估计得出的估计值的折衷。很清楚,“借力”的想法是元分析(meta analysis)的一种变形。

1.4 简 史

在本节中,我们将针对多层次模型的历史做一简要回顾。本节从 De Leeuw 与 Kreft(1986)的介绍再加以延伸,将不同领域的一些重要论文加以整理,探究彼此的关系。我们特别将重点放在综述文献与教科书的整理上,目的是凸显出在教育统计与其他领域中多层次分析的发展脉络是非常类似的。此外,本节也对不同领域的文献进行概论性质的介绍。

在某一个领域所发展的工具往往会被其他领域所引用。统计学作为一个学术领域,其中一个功能即是作为各学科之间的整合者,提供数据分析的纪录。通过本节的整理,我们可以发现多层次模型分析是汇聚了各种已被提出的模型概念,例如方差成分模型(1.4.1)、区块数据分析经济学的随机系数模型(1.4.2)、变动系数模型(1.4.3 与 1.4.4)、追踪的数据分析(1.4.5)、成长曲线模型(1.4.6),以及贝叶斯与经验贝叶斯估计法(1.4.7)。至于调节变量(1.4.8)与“斜率结果”(1.4.9)的概念也与多层次模型有关。

1.4.1 方差成分

方差成分分析(variance component analysis)(或混合模型分析;mixed model analysis)具有很长及复杂的历史,在 Searle 等人(1992)的书中有详细的介绍。方差成分分析最早是由 Airy(1861)应用于天文学中,进一步的发展则由 Fisher(1918; 1925)完成。有关固定效果与随机效果的区分,以及混合模型的诞生,则是 Eisenhart(1947)的贡献。在 1950 到 1970 年间,这一个领域则由 Henderson(1953)方差成分估计法所主导。在 1970 年左右,计算机革命使得最大似然估计得以盛行(参见 Harley & Rao,1967; Hemmerle & Hartley,1973; Harville,1977; Searle,1979; Thomson,1980)。从 1970 年开始,大量的讨论集中于计算过程,这时期的研究让研究者对于为何需要进行方差分析(ANOVA)有更深一层的理解,建议参考 Engel(1990)的讨论。另外 Speed(1987)与 Samuels 等人(1991)的论文也值得参考。

1.4.2 随机系数

随机系数模型(random coefficient model)是由经济计量领域的 Wald(1947)与 Rubin(1950)所提出,计算机革命使得此种模型得以成为实际可行的技术。在 1970 年代,Rosenberg(1973)、Spj∅tvoll(1977)与 Swamy(1971)等人的专文对此一模型做了详细的说明。Johnson(1977,1980)出版了详细的操作手册,最近的发展趋势则是企图将随机系数模型以准参数(semiparametric)形式来操作,也就是随机效果的分配不是假设为常态的情况,而由数据中直接加以估计。有关线性的状况,可参考 Beran 与 Hall(1992)的书,有关非线性的状况,则可参考 Davidian 与 Gallant(1992)的著作。

1.4.3 变动系数

变动系数模型(variable coefficient model)是一种一般化的模型。每一组有自己的一套回归系数,通过非线性的方式来计算得出。基本上,此种模型的估计是由修匀程序等计算法则来进行。最近的文献可以参考 Hastie 与 Tibshirani(1993)的著作,该著作也讨论了一般化的加成模型(Hastie 与 Tibshirani,1990)。这些模型中的系数是固定的。

1.4.4 变化系数

在随机系数模型中,结果变量与解释变量的关系是以线性回归模型来处理,但是在特定的时间点 t 之下,有一组回归系数,此时不同时间点的回归系数为变化系数(changing coefficient)。为了不让估计参数过度膨胀,我们可以假设有一个第二阶层模型来处理不同时间下的回归系数,通常会假设回归系数带有自我回归的路径模型(autoregressive path model)。最近对于此一模型的关注颇多,Chow(1984)有一篇论文综述至 1984 年的文献,而 Kalman 的控制系统理论与此一模型有着密切的关系。

1.4.5 追踪数据

在经济学领域,至少在微观经济学,一群个体在不同的时间下进行观察的追踪数据(panel data)是非常受到重视的一种数据型态。我们推荐 Chamberlain(1984)的论文以及 Hsiao(1986)的著作。这种模型通常是变动系数回归模型,有时则是随机系数模型。在许多情况下,他们即是混合模型或变异成分模型(Wansbeek,1980)。

1.4.6 成长曲线与重复量数

成长曲线模型(growth curve model)是由生物计量领域所研究的课题。关键文

献是 Pothoff 与 Roy(1964)的论文,文中他们利用多变量线性模型来处理一个平衡的两阶层模型,而在第二层没有随机成分。Rao(1965)结合了成长曲线与随机系数模型。另一方面,利用 MANOVA 来处理成长曲线模型,并延伸到重复量数模型的讨论,则可参见 Geisser(1980)与 Timm(1980)的著作。有关成长曲线模型与多层次模型之间关系的详细讨论,可参见 Strenio et al.(1983)与 Jennrich & Schluchter(1986)的著作。

1.4.7 贝叶斯线性模型与经验贝叶斯估计

多层次模型与线性模型贝叶斯分析之间具有非常密切的正式关系,在 Lindley, Smith, Leamer, Zellner 与一些其他学者作品中有详细讨论(Lindley & Smith, 1972; Smith, 1973)。我们会说他们具有“正式”的关系,是因为贝叶斯方法就是假定系数为随机。若以频率学派的混合模型(frequentist mixture models)来解释多层次模型,可以解释得一样好。

利用 James-Stein 的理论,线性模型中的缩动估计值(shrinkage estimator),也可以以误差均方的形式来考虑。Efron 与 Morris(1975)和 Morris(1983)的经典论文诠释了缩动估计的数据分析的层次。国家研究委员会(NRC)(1992)讨论了“借力”的名词,并有详细的说明。该报告虽专注于元分析并据以作为主要的应用范围,但文中有关方法学的讨论也具有参考价值。

1.4.8 调节变量

调节变量(moderator variable)的概念并不容易定义,Baron 与 Kenny(1986)的著作有相当深入的综述。Velicer(1972)从不同组别下的不同回归模型的角度来介绍此概念。更早的一篇论文,Saunders(1956)则明示可将回归方程式中的系数作为第二组回归方程式的反应变量。

1.4.9 斜率结果

“斜率结果”分析(‘slope-as-outcome’ analysis)是由 Burstein 等人(1978)所提出来代替 Cronbach 的方差分解技术。Burstein 等人(1989)的著作是很好的历史回顾。该取向的技术其实是两阶段常规最小二乘法回归分析(ordinary least squares; OLS),但是其背后的计算原理并不是很清楚。也就是说,随机系数模型是斜率结果取向的严谨版本。

1.5 进一步的读物

据我们所知，有三本有关随机系数回归或多层次分析的书，都比本书谈得更为深入。读者若具有深厚的统计、矩阵代数基础，这些书是非常理想的入门教材。以下分别就这三本书的大略内容进行介绍。

Bryk 与 Raudenbush(1992)是从教育统计学的角度切入，适用于教育界或学校研究。书中所介绍的多层次材料利用 HLM 软件来做示范说明，利用该软件的界面来介绍多层次分析的一些变化。书中的例子多为教育领域，因此变量的数目较少，对于模型的评鉴与替代模型的使用则讨论得较少。书中有专章讨论元分析的应用，这种应用是在方差与协方差成分是假设为已知的情况下进行，同时也有章节介绍重复量数与成长曲线的应用。整体来说，这本书对于多层次分析这种新技术介绍得非常详尽清楚，在对教育与社会学领域的研究问题的有效解决方面向前迈出了重要的一步[6]。

Longford(1993)的书比起 Bryk 与 Raudenbush(1992)的书偏向技术层面，而较不专注于多层次模型与“斜率结果”策略。就如同书名(*Random Coefficient Models*)所言，该书是以随机系数模型这种特殊的混合线性模型为主，模型中的方差与协方差成分是从随机回归系数(包括截距)中获得。书中对于随机系数模型的处理说明得非常完整清楚，以随机协方差分析为起点，并延伸到类别与多变量的结果变量的应用。

Goldstein(1987,1995)的书则有两个版本，最近一版的书是 1987 年版的延伸与扩充，因此可以完全取代旧的版本。该书针对多层次模型的基本原则详加说明，并进一步把这些原则应用到一些基本的统计模型的分析步骤中。虽然细节谈得不多，但是所延伸出来的重要概念则反复加以强调。书中的章节包括了类别反应变量、事件历史分析、变量中的误差与联立方程式模型、广义线性模型，等等。Goldstein 的书并非是一本概论性的书，而是针对一组模型详加解释的进阶性专著，甚至比较像做研究的指南，报告了这些模型截至目前的发展，也试图描绘未来的应用前景。

---

[6] Bryk 与 Raudenbush(1992)的著作是以他们所发展的 HLM 软件作为论述的基础，对于 HLM 软件的熟悉与运用，阅读 Bryk 与 Raudenbush(1992)的著作十分有帮助。——译者注

1.6 软 件

目前有多种软件可以用来分析线性模型中的阶层数据。这些模型可以说是混合线性模型(mixed linear model)的特殊情况,因此可以利用混合线性模型的原理来分析数据[7]。另一方面,一些特殊的混合线性模型则必须利用特殊的软件来分析。因此,在前面的篇幅中,先简略介绍一些一般性的统计软件,并指出优缺点,其他更进一步说明软件的应用,可参考 Kreft 等人(1994)与 Kreft 等人(1990)的介绍。

本书将以 Rasbash 等人(1991)在英国所发展的 MLn 软件来进行示范,此举并不是说 MLn 软件优于其他软件(虽然有人如此认为),主要在于 MLn 使用传统的符号系统来描述数据与模型。同时此一符号系统与数学方程式的使用方式相同,有助于不善于处理数学方程式的读者来研习。然而本书也不是 MLn 的操作手册,有关 MLn 的详细介绍与说明可参考 Woodhouse(1995), Rashash & Woodhouse (1995)与 Prosser et al. (1991)的著作。

1.6.1 HLM

在教育统计中,HLM 软件扮演重要的角色。在美国,HLM 软件已经被正式出版,成为多层次分析的正式软件[8],其操作手册非常详细,软件的操作接口简单,没有太多需要设定之处。软件的运作模式遵循典型的斜率结果,对于教育研究者来说可以说是非常熟悉。软件的操作接口非常简单,是一些简单的问答,同时软件也提供有用的统计检验与描述统计的数据,进一步的信息可以参考下列网址:

http://www.ssicentral.com/hlm/index.html

1.6.2 VARCL

VARCL 软件由 Longford(1990)所开发,可应用于阶层数据的方差成分分析。VARCL 是设计给随机系数分析之用,并非专门写给多层次分析的软件,因此使用

---

7 在 SPSS 第 12 版之后,新增的混合模型模块(mixed models)可进行多层次模型分析,包括阶层线性模型与随机系数模型。亦即在本书所介绍的各种软件之外,SPSS 也是一种可处理本书所提及的各种模型的应用软件。——译者注

8 HLM 的第六版已经在 2005 年出版,并有学生版供自由下载,第六版的使用者接口与绘图功能更加完善,有关软件的详细资料、指导手册与学生版软件下载请参考下列网站:http://www.ssicentral.com。台湾有统计方法学学会举办研习工作坊,相关信息请参考下列网站:http://www.tasm-sem.org。——译者注

者必须自己决定各变量的系数的固定与随机部分,无法以简单的设定来检验跨层级的交互作用。VARCL 使用一般的量尺法,延伸到波松与二项分布反应形式。VARCL 有两种版本,第一版分析最多到三阶层的随机斜率与截距,第二版则可处理最高达九个阶层的随机截距与斜率。进一步的信息可参考网址:

http://www.gamma.rug.nl/iechome.html

1.6.3 BMDP5-V

BMDP5-V 是著名的 BMDP 软件的一部分,由 Jennrich 与 Schluchter(1986)所开发,这套软件的目的在于分析重复量数的数据,也就是说,适用于第二层各分析单位下具有少数平衡观察数据(重复个数相同)。它也可以应用于分析多层次的数据,但是必须撰写复杂的语法指令。BMDP5-V 可以选择不同的算法技术,也可以描绘出第二层的残差分布状态。有关的信息可参考下列网址:

http://www.spss.com/software/science/bmdp/

1.6.4 MLn

MLn 是由伦敦大学教育学院的多层次项目(multilevel project)所开发的软件。MLn 的一项特别有用的功能是整合了一般统计分析的 NANOSTAT 软件,使得一些基本的统计运算与检定可以一并完成。MLn 可以分析多达 15 层的跨层级嵌套结构数据。它也具有宏语言的功能,可以用来开发结果变量为类别变量的模块。有关 MLn 的进一步信息可以参考下列网站:

http://www.ioe.ac.uk/multilevel/

1.6.5 PROC MIXED

PROC MIXED 是 SAS 软件当中的混合模型分析模块,此一模块与 BMDP5-V 版本软件兼容。进一步的信息可参考下列网址:

http://www.sas.com/

1.6.6 MIXOR and MIXREG

MIXOR 与 MIXREG 由 Don Hedeker 所开发(Hedeker & Gibbobs, 1993a, 1993b)。Hedeker 在其博士论文(Hedeker, 1989)和 Hedeker 与 Gibbons(1994)的论文中说明了理论基础。MIXOR 可以处理结果变量为顺序变量的多层次分析,MIXREG 则可处理具有自我相关误差项的多层次分析。这两个软件都有 PC 版与 Macintosh 版本,手册与软件可自下列网站下载,此一软件使用的是 Newton-Raphson 算法来处理两层次的模型。相关信息可参考下列网址:

http://www.uic.edu/~hedeker/mixdos.html

1.7 摘 要

本章作为本书的绪论，介绍了有关多层次分析的各种重要概念。文中说明了一些旧概念被赋予了新意义，例如脉络模型与微观/宏观层次的概念，对于一些更细节的概念，例如随机与固定系数，以及组内相关等也予以简单介绍。此外，我们利用一些真实研究的范例说明了多层次模型的用途，特别是应用于复杂脉络下，人类行为的新理论与新观点的检验，例如将跨层级的交互作用纳入模型中。文中还简述了多层次分析的历史，显示了此一技术在各领域都可应用，以解决不同的研究课题。最后，文中介绍了分析软件最近的发展，显示多层次技术已经越来越普及，操作上也更为便捷容易。

批 注

1. 组内相关是指组间方差占全体方差的比例。

脉络模型概述 2

2.1 绪 论

本章我们将讨论传统的回归模型在方差分解上所造成的不同。整体回归模型(total regression model)仅允许个体变异存在,而聚合模型(aggregated model)恰好相反,仅允许脉络间(宏观层次)存在变异。传统的脉络模型、协方差模型、Cronbach 模型与不同型态的多层次模型虽然都将数据的方差分解成脉络间与脉络内两个部分,但是各有不同的做法。为了解释这些模型的差异,本章使用美国全国教育纵贯研究(1988)(National Education Longitudinal Study; NELS88)的一部分数据分析结果为范例[1]。

2.2 模 型

在本章当中,我们将讨论一系列传统的线性模型与 OLS 回归模型的变化运用。从整体回归(total or pooled regression)——完全忽视组间变异,到聚合回归(aggregate regression)——完全忽略组内变异。若从另一个向度来看,模型的变化是从每一组个别进行回归可以得到多组回归系数变化到只有一组参数的单一回归。

在许多情况下,把宏观的结构特性更清楚地纳入模型有其必要,若我们把微观

---

[1] 在台湾也有两个类似的数据库正在执行中,其一为"台湾教育长期追踪数据库"(Taiwan Education Panel Survey,简称 TEPS)。这个数据库从 2001 年开始,对当年为中学一年级以及高中、高职和五专二年级之学生、学生家长、老师和学校,进行为期七年二至四次的搜集数据。更多信息请参考网址http://www.teps.sinica.edu.tw。第二个数据库为"台湾高等教育数据库",自 2003 年 7 月起开始,搜集有关台湾高教院校特质与状况、各科系在校学生素质、各学科毕业生、课程、教职员、经费以及校园环境设备等数据的全台湾性的数据库,并供研究人员做分析高教问题之用。更多信息请参考网址http://www.cher.nthu.edu.tw。——译者注

层次与分组层次的解释变量同时纳入回归模型,称为脉络分析(contextual analysis)。在脉络分析中,组别数据是不能被忽略的信息,此时观察单位为组别,显示研究者的兴趣不仅在个体,更是在脉络层次。

从脉络分析的技术层面来看,线性模型的自由参数是用下列的模型来估计:y 为反应变量,x 是微观层次解释变量,z 是脉络层次解释变量(注 1)。下标 i 表示个体,j 表示脉络。模型如下:

$$y_{ij} = a + bx_{ij} + cz_j + \varepsilon_{ij} \tag{2.1}$$

ε_{ij} 为扰动项(disturbance),具有中心化、等分散性、独立性等特征。ε_{ij} 的分配的期望值为 0,方差为 σ^2。一般来说,两层都有可能具有一个以上的解释变量。公式 2.1 可以另一种可展现结构特性的不同形式来表示:

$$y_{ij} = a_j + bx_{ij} + \varepsilon_{ij} \tag{2.2a}$$

$$a_j = a + cz_j \tag{2.2b}$$

公式 2.2b 显示公式 2.1 的脉络模型是一个变动截距模型(varing intercept model),也就是每一个组的回归模型都具有相同的斜率,因此有相同的误差方差 σ^2,但是各组截距不同。本章所讨论的各种模型,差别在于变动截距与分组层次解释变量间关系的不同设定。

2.3　资　料

在本章当中,我们将举出几个以 NELS88 数据库来进行分析的研究范例。此数据库是由美国教育部国家教育统计中心所搜集的长期纵贯资料。纵贯资料的第一波测量是八年级的学生,测量范围涵盖将近 1 000 所学校(800 所公立、200 所私立),第一年(1988)将近有 20 000 名学生参与调查。在 1988 年时,美国全国超过 38 000 所学校的八年级学生总数约有三百万人,这个样本具有总体的代表性。本章所使用的数据是从 NELS88 数据库所选取的部分数据,以作为示范说明之用,在后续的章节中,我们则会使用全部的数据来进行分析,讨论较为实际的分析问题。

我们所选取的数据库是在 NELS88 的 1 003 所学校中以人工方式选择了 10 所,包含 260 位学生,为全部资料 21 580 学生当中的小部分。模型中所使用的变量为家庭作业时数(为微观层次变量)与数学成绩;前者将作为我们的解释变量,后者作为我们的反应变量。因此,我们希望能够发现做作业的时数[家庭作业]会导致或能预测[数学成绩](注 2)的高低。我们所关心的学校层次特征为公私立之别(以[公立]表示),每一个学校可被归为某一种类型,公立学校编为 1,私立学校编

为0。我们对于学校的选择,是考虑数学成绩与做作业时间具有高度正向关系与高度负向关系的两个极端,如此选择的原因是想让我们的回归线显现出明显的变异。

在NELS88的数据库中,大多数学校的做作业时间与数学成绩之间都具有正向的关系。表2.1与表2.2列出了这10所学校的状况。表2.1列出了10所学校的平均数学成绩(答对题数)与做作业时间(每周几个小时);表2.2则列出了方差、协方差与相关系数。有关变量的详细内容列于附录。在针对不同模型的分析结果进行讨论之前,我们先把组间(或脉络间)与组内(或脉络内)的参数估计的差别再做一次详细说明。

表2.1 NELS88数据库中10所被取样的学校的各校平均数

学 校	组样本规模	平均数学成绩	平均做作业时数
1	23	45.8	1.39
2	20	42.2	2.35
3	24	53.2	1.83
4	22	43.6	1.64
5	22	49.7	0.86
6	20	46.4	1.15
7	67	62.8	3.30
8	21	49.6	2.10
9	21	46.3	1.33
10	20	47.8	1.60

表2.2 NELS88中10所被取样学校的各校两变量离散性与相关

学 校	方差/协方差		相 关
A	55.20	-4.24	-0.52
	-4.24	1.19	
B	65.10	-4.65	-0.45
	-4.65	1.63	
C	126.30	9.62	0.77
	9.62	1.22	
D	94.10	11.9	0.84
	11.90	2.14	
E	69.20	-2.71	-0.43
	-2.71	0.57	
F	17.00	-1.56	-0.48
	-1.56	0.63	
G	31.20	3.24	0.34
	3.24	2.92	
H	101.10	7.94	0.71

续表

学 校	方差/协方差		相 关
	7.94	1.22	
I	86.60	4.61	0.56
	4.61	0.79	
J	120.90	12.30	0.80
	12.30	1.94	

2.4 方差分解

在两层结构的阶层嵌套数据中，观察变量的方差与协方差可以被切割成组间与组内矩阵两个部分。变量的组间与组内变异的切割并不全然是直接的，每一种分析技术之下又有所不同。为了说明不同模型的回归系数的定义，我们使用相关比的符号来说明。相关比(correlation ratio；或以 η^2 表示)是指组间方差占一个变量总方差的比例，解释如下。本书稍早所使用的公式，例如 x 为解释变量，y 为反应变量，可以切割成组间与组内两个部分。以变异切割关系式表示如下：

$$V_T(x) = V_B(x) + V_W(x) \tag{2.3a}$$

$$V_T(y) = V_B(y) + V_W(y) \tag{2.3b}$$

其中下标 T,B,W 分别表示总变异、组间变异、组内变异。如果熟悉 ANCOVA 的符号，变量 x 与 y 的总协方差也可以以相同的方式切割成组间与组内两个部分，其中 C 表示协方差：

$$C_T(x,y) = C_B(x,y) + C_W(x,y) \tag{2.3c}$$

以所有样本计算出来的回归系数为 b_T，以组为观察单位所计算出来的回归系数为 b_B，以组内数据所计算出来的回归系数为 b_W，这三者可以利用组间或组内相对于全体变异的角度来定义(注 3)：

$$b_T \triangleq \frac{C_T(x,y)}{V_T(x)} \tag{2.4a}$$

$$b_B \triangleq \frac{C_B(x,y)}{V_B(x)} \tag{2.4b}$$

$$b_W \triangleq \frac{C_W(x,y)}{V_W(x)} \tag{2.4c}$$

上述三个系数与相关比(η^2)有关，x 与 y 变量相关比的关系式分别如下：

$$\eta^2(x) \triangleq \frac{V_B(x)}{V_T(x)} \tag{2.5a}$$

$$\eta^2(y) \triangleq \frac{V_B(y)}{V_T(y)} \tag{2.5b}$$

上述公式显示,反应变量的组间变异是 y 变量在组间的变异占总变异的比例,亦即与第一章所介绍的组内相关的定义一样。同时,

$$1 - \eta^2(x) = \frac{V_W(x)}{V_T(x)} \tag{2.6a}$$

$$1 - \eta^2(y) = \frac{V_W(y)}{V_T(y)} \tag{2.6b}$$

组内变异的百分比是 $1 - \eta^2(x)$,等于组内变异与总变异的比值。从一般的回归分析可知,对于整体样本的回归,系数 b 的最佳估计值是 b_T ,此一估计不涉及组间的比较问题。b_T 是组间回归系数 b_B 与组内回归系数 b_W 的加权组合的结果,如下列公式所示:

$$b_T = \eta^2(x)b_B + (1 - \eta^2(x))b_W \tag{2.7}$$

最早使用公式 2.7 来进行脉络分析的是 Duncan 等人(1966),之后 Boyd 与 Iversen(1979)与 Burstein(1980)加以沿用。

一个整体模型(未将脉络效果从个体效果中加以分离的模型),只有在 $b_T = b_W$ 且 $b_B = 0$,或是 $\eta^2(x) = 0$ 的情况下才会得到可信赖的估计值,也就是没有脉络效果的情况下。从公式 2.7 可以很清楚地看出,在没有脉络效果的情况下($\eta^2(x) = 0$), b_T 才是个体层次斜率的有效估计。没有脉络效果意味着 x 与 y 的关系在所有的脉络下是一样的,没有组间的差异。

当 $b_T = b_B$,也就是 $b_W = 0$ 或 $\eta^2(x) = 1$ 时,情况则与上面相反,回归的效果完全是脉络间的差异(脉络效果),并没有任何的个体效果。x 与 y 的关系对于同一个脉络下的每一个个体是一样的。在很多情况下,变量在各层次都有测量数据,而 b_T 是 b_B 与 b_W 的加权组合,权数分别为 $\eta^2(x)$ 与 $1 - \eta^2(x)$ 。如果一个研究的目的是发现脉络效果,那么整体模型就不是一个理想的模型。在我们的范例中,脉络效果是学校类型(公私立)的差异,每一个学校有不同的数学成绩表现的解释型态,以整体模型来分析并不适当。

2.5　整体回归

首先,我们所讨论的技术相对比较简单,既非多层次分析,甚至不是脉络性的分析。我们如果把所有的样本放在一起,求出单一条回归线来描述家庭作业时数对于数学成绩的影响,此时并没有使用到任何有关学校的信息。也就是某一群学生来自同一个学校,另外一群学生来自另一个学校的这种样本结构特性并没有反映在模型当中。

对于全部的个别观察值进行回归分析,忽略其组别特性,就如同我们拿掉公式2.1中的下标 j,模型变成下述形式:

$$\underline{y}_{ij} = a + bx_{ij} + \underline{\varepsilon}_{ij} \tag{2.8}$$

其中 ε_{ij} 为独立的随机变量,平均数为0,方差为常数 σ^2 。为了完整说明,并为了后续的比较,我们可以提出一个相对应的模型:虚无模型(null model),没有任何解释变量,仅带有截距 a :

$$\underline{y}_{ij} = a + \underline{\varepsilon}_{ij} \tag{2.9}$$

对于所有的个别观察值所进行的回归分析称为整体回归(total regression),此时个体为分析的单位,也是抽样与做决策的单位。这样的分析意味着我们并不预期学校的差别对于数学成绩会有系统性的影响,所有关于学校的影响是囊括在误差项之内。而个体嵌套于组别之中的此一抽样特性是被忽略的事实,且被假定对于研究目的没有什么重要性的影响。公式2.2a的脉络模型的截距 a_j ,在整体回归中被假设对 j 个组都是相同的数值。分析结果列于表2.3,从表中我们可以看到每增加一小时的做作业时间,将会增加3.6分的数学预测分数。

表2.3　10个学校的整体回归分析结果[译者分析1]

	虚无模型		带有[家庭作业]模型	
	估计值	标准误	估计值	标准误
截距	51.3	0.69	44.1	0.98
斜率 b_T	N/A[1]		3.6	0.39
R^2	0.00		0.25	
$\hat{\sigma}$	11.10		9.60	

1　为英文的 none available 的缩写,表示无法提供之意。——译者注

2.6 聚合回归

另一个比较原始、未加工的方法,可以将脉络因素(学生来自的校别)在回归中加以考虑,是取用学校的平均值来进行回归分析,称为聚合分析(aggregate analysis)。基本上,我们并没有什么理由认为整体回归所得到的回归系数,与以聚合数据来进行回归分析得到的回归系数会相同。事实上,我们可以很容易地通过例子来显示这两种方法得到的结果是非常不一样。

若 10 个学校的家庭作业平均数为 $x_{.j}$,数学成绩平均数为 $y_{.j}$,可以得到下列模型:

$$y_{.j} = a + bx_{.j} + \varepsilon_j \tag{2.10}$$

公式中下标的句点"."取代了代表个体的 i,表示现在的 x 与 y 变量是将个体的数据加总求平均的形式。同样的,我们假设 ε_{ij} 的平均数为 0,但方差为 $n_j^{-1}\sigma^2$,与整体回归相比,多了一项 n_j^{-1} ,是因为误差方差是 j 个组(n_j)的残差(每一组的方差为 σ^2)的平均值。此时的回归模型是加权回归,权数等于 n_j ,此时的回归是异分散性(heteroscedastic),因为每一组的误差变异并不相等。

很明显,聚合回归忽略了所有组内(学校内)的变异,因此流失了大量可能是很重要的变异。

表 2.4 列出了(加权)聚合回归的结果,我们可以看到把组内变异排除之后, R^2 (多元相关平方)大幅提高,但是标准误也扩大很多,因为这个回归分析是基于 10 个学校的观察单位的结果。

表 2.4 10 个学校的聚合回归分析结果[译者分析2]

	虚无模型		带有[家庭作业]模型	
	估计值	标准误	估计值	标准误
截距	51.3	2.44	37.1	4.03
斜率 b_B	N/A		7.0	1.84
R^2	0.00		0.64	
$\hat{\sigma}$	39.30		24.9	

聚合回归的解释必须十分谨慎,从预测的观点来看,我们仅能说,平均而言,若 A 校学生比 B 校学生每多做一个小时的作业,A 校数学成绩会比 B 校多 7 分。此一说法并不涉及任何有关学生个体成绩的预测,但是如果以聚合回归的结果来进

行学生的成绩预测,即是一种区位谬误或生态谬误(ecological fallacy)(Robinsion,1950)。

2.7 脉络模型

脉络模型(contextual model)在过去已经被大量使用于探讨团体效果对于个体行为的影响的研究之中。典型的分析方法,是把个体层次的变项取组平均数后,作为脉络变量(contextual variable)。例如学校的特征为学生们的平均做作业时间 $x_{.j}$,个别学生做作业时间则为 x_{ij}。在同一个回归模型中,相同的测量被使用两次,第一次是原来学生做作业时间,第二次是计算出每校的平均时间。家庭作业的平均数是从每一个学校聚合而得。此时模型如下:

$$\underline{y}_{ij} = a + bx_{ij} + \varepsilon_{ij} \tag{2.11a}$$

$$a_j = a + cx_{.j} \tag{2.11b}$$

整理之后,可以得到脉络模型的整体方程式如下:

$$\underline{y}_{ij} = a + bx_{ij} + cx_{.j} + \varepsilon_{ij} \tag{2.12}$$

脉络模型分析的结果列于表2.5。由表2.5的结果显示,公式2.12中的 b 系数最佳估计值是 b_W,c 的最佳估计值是 $b_B - b_W$,有关更详细的讨论,可参考 Duncan 等人(1966)、Boyd & Iversen(1979)与 Burstein(1980)有关脉络模型中的组内回归系数 b_W 与组间回归系数 b_B 的混淆问题的讨论。

表2.5 10个学校的脉络模型分析结果[译者分析3]

	虚无模型		带有[家庭作业]模型	
	估计值	标准误	估计值	标准误
截距	51.3	0.69	37.1	1.46
斜率 b_W	N/A		2.1	0.43
脉络效果 b_B	N/A		4.9	0.79
R^2	0.00		0.34	
$\hat{\sigma}$	11.1		9.0	

脉络模型存在一些技术上的问题。其一是多元共线性,另一个问题是分析的层次。多元共线性(multicollinearity)因为个体解释变量[家庭作业]与分组层次的家庭作业平均数的相关而存在。另一方面,分析层次的问题在于,由于反应变量是个体层次的测量数据,因此分析层次实际是针对个体层次而为。聚合变量基本上

仍是个体层次的测量,因此仍然只是针对其中一个层次进行回归分析而忽略数据的阶层嵌套结构。从数据可知,脉络模型当中的脉络效果是 b_B 与 b_W 的差,亦即表2.4 的 b_B (7.0)减去表2.5 的 b_W (2.1)所得到的4.9,此结果在表2.5 中也可以看到。很明显,个体与分组效果在 c 参数上是有所混淆的,导致有关分析结果与关系的解释可能受到扭曲。

2.8 Cronbach 模型

Cronbach 模型(Cronbach & Webb, 1975)对于个体效果与团体平均效果两种效果一起对反应变量产生影响的分析,提出一个较为清楚的图像。个体层次的解释变量事先经过组平均中心化(组平减)处理,也就是将各组观察值的数值减去组平均数,如公式2.13 所示:

$$y_{ij} = a + b_1(x_{ij} - x_{.j}) + b_2(x_{.j} - x_{..}) + \varepsilon_{ij} \tag{2.13}$$

公式2.13 当中,经过组平减的个体分数 $x_{ij} - x_{.j}$ 所形成的变量与组平减的分组层次分数 $x_{.j} - x_{..}$ 为正交关系。以我们的例子来说,10 个学校的组平减分数 $x_{ij} - x_{.j}$ 为某一所学校当中每一个学生做家庭作业的时间减去这个学校的平均时间而得到。换句话说,原始分数以组平均数为中心转换成离差分数(deviation score)。模型当中的这个中心化解释变量提供了一个让两个层次同时以[家庭作业]这个变量分别得到两个解释变量以避免他们具有相关的简便方法。在 Cronbach 模型中,用于回归分析的两个预测变量是中心化(组平减)的个体层次解释变量[家庭作业]与中心化的家庭作业组平均数,结果列于表2.6。由于两个预测变量呈现直交,b_1 的最佳估计值为 b_W,也等于前一节脉络模型的估计值。

表2.6 10 个学校的 Cronbach 模型分析结果[译者分析4]

	虚无模型		带有[家庭作业]模型	
	估计值	标准误	估计值	标准误
截距	51.3	0.69	37.1	1.46
斜率 b_B	N/A		2.1	0.43
脉络效果 $b_B - b_W$	N/A		7.0	0.67
R^2	0.00		0.34	
$\hat{\sigma}$	11.1		9.0	

Cronbach 模型与脉络模型的差异在于脉络效果的估计。在 Cronbach 模型中,b_2 等于 b_B,也就是聚合模型 b_B 的效果。因此在 Cronbach 模型中,组间与组内效果

不会有所混淆。

虽然 Cronbach 模型解决了个体层次与聚合变量之间存在相关的共线性问题，但参数的显著性检验仍然像脉络模型一样具有疑义。在这两种脉络模型中，分析都是以低阶层数据来处理，导致团体平均数参数的标准误都被低估，造成显著性检验的第一类型错误膨胀。团体平均数的数目与组数相同。我们的数据库中有 10 所学校，每所有 22 个观察值，因此总观察值为 220 个，分析中的标准误是以这 220 个观察值为基础所计算得到，而不是真正的 10 所学校。另一方面，脉络模型中还有另一个因素——组内相关，也会影响到标准误的正确性。在第一章我们已经讨论过，当组内相关越高，也会提高 α 水平。在第 5.2 节有关中心化的问题讨论中，我们还会再就 Cronbach 模型的特性加以说明。

2.9 协方差分析

协方差分析（analysis of covariance; ANCOVA）是另一种可以用来分析分组数据的一种统计技术。ANCOVA 模型同时包含学校与学生层次的分析，但是角色各有不同。个体层次解释变量涉及回归分析，学校则影响截距。换言之，ANCOVA 同时纳入量化与质化变量于模型之中，因而具有混合模型的特性，可以说是带有虚拟变量的回归分析，虚拟变量的使用可以将团体特性纳入回归模型。回归模型可以分析量化因子（例如学生做作业的时间）的影响，ANCOVA 模型则可以把质性因子（例如学校差异）纳入模型。

ANCOVA 可以应用于不同目的的脉络分析之上。ANCOVA 可以调整既存组间差异之后来探讨组别的效果，例如把各校学生在做作业时间上的差异进行调整（控制）之后，检验各学校的数学成绩有无差异。这样的分析让我们可以得知学校的平均成绩是否不同，以及哪一个学校平均来说表现较佳。在 ANCOVA 中，个体效果是被忽略的，或被视为是干扰，所强调的是组间的效果。

在 ANCOVA 中，个体解释变量的功能是协变量（covariate），而分组变量则是分析的重点。由于这种设计源自于实验研究，在 ANCOVA 中的分组通常是指不同的实验处理。协方差分析模型如下：

$$\underline{y}_{ij} = a_j + bx_{ij} + \varepsilon_{ij} \tag{2.14}$$

公式 2.14 中，截距 a_j 的不同数值表示各校在数学成绩上有不同的起始值（starting value），有的学校较高，有的学校较低。ANCOVA 的基本假设是各校都有相同的斜率（模型中的 b 系数），表示我们假设各校的家庭作业时间与数学成绩表现之间的关系强度相同。以 NELS88 的数据为范例所得到的结果列于表 2.7。

我们允许不同的学校有不同的起始值(不同的截距 a),然后假设各校每增加一单位做作业时间所增加的数学成绩幅度是相同的,公式 2.14 与公式 2.2a 相同,但是公式 2.2b 则消失。也就是说,对于截距 a_j 并没有额外结构上的限定,它可以为任何数值。

因为 ANCOVA 是比较 k 组之间的差异,自由度皆为 $k-1$,因此模型提供了可以归因于脉络的整体差异方差的上限。

表 2.7 10 个学校的协方差分析结果[译者分析5]

	虚无模型		带有[家庭作业]模型	
	估计值	标准误	估计值	标准误
A	45.8	1.77	42.8	1.75
B	42.2	1.90	37.1	2.00
C	53.2	1.73	49.3	1.78
D	43.6	1.81	40.1	1.82
E	49.7	1.81	47.9	1.74
F	46.4	1.90	44.0	1.84
G	62.8	1.03	55.7	1.60
H	49.6	1.85	45.1	1.92
I	46.3	1.85	43.5	1.82
J	47.8	1.90	44.4	1.89
斜率	N/A		2.1	0.38
R^2	0.44		0.50	
$\hat{\sigma}$	8.5		8.0	

相对于传统的脉络模型(公式 2.1),ANCOVA 并无法告诉我们哪些脉络(学校)的特性可以解释脉络间的差异,它所能提供的信息是组间的效果(整体而言)有多大,这是通过截距项的组间被解释变异(explained variance)指标所反应。

ANCOVA 最重要的优势是比传统的脉络模型(公式 2.2a 与 2.2b)有较高的预测检定力。ANCOVA 把脉络平均数之间的变异全部纳入分析,其中包含了脉络模型中与脉络有关的解释变量的解释变异。但是脉络模型的一个强处是可以辨识出影响组间差异的脉络特性。多数研究者认为,在进行脉络分析之前,可利用 ANOVA 或 ANCOVA 检验整个组间效果的强度。基本上,如果截距 a_j 只能提供有限的解释变异,在后续的模型中,是不会有什么脉络特性可以用来解释反应变量的变异量。但是,值得注意的是,这个说法是当脉络间变异是从截距的部分(主要效果)来加以研究时才成立。越来越多的研究关心解释变量与反应变量间关系的脉络差异,也就是模型 2.14 中的 b 系数。ANCOVA 假设 k 个解释变量(或协方差)的每一个对于反应变量在所有的脉络下都有一样的影响力,这是不切实际的假设,我们在

下一章还会加以说明。每一个学校都可能需要发展一个自己的解释模型,有自己独特的解释变量(家庭作业时数)与反应变量(数学成绩)的关系模式。

2.10 脉络模型的 MLn 分析

假设我们今天有一个数据文件 schools. dat,其中有四个变量,第一个是学生的特征,第二是学校的身份变量,第三是学生花费在数学作业的时间,第四是学生的数学成绩。文件中每一个学生有一笔数据,四个变量之间以空格隔开。

我们可以开始 MLn 分析。首先会给我们一个空白的工作窗口,在工作表中,我们设定计算机读取四个变量,放入四个变量字段中:

```
DINPUT C1-C4
```

计算机软件将会出现一个光标要求输入文件名称,我们告诉计算机文件名是 schools. dat,计算机就会把四个变量的数据读入,然后我们必须各给一个名字:

```
NAME C1 'student'
NAME C2 'school'
NAME C3 'homew'
NAME C4 'math'
```

为了进行脉络分析,我们需要一些其他变量。一个常数项,家庭作业与数学成绩的各校平均数,做作业时间与数学成绩我们以组平减来处理。此时我们必须先给定这些字段的名称:

```
NAME C5 'cons'
NAME C6 'meanhomew'
NAME C7 'meanmath'
NAME C8 'devhomew'
NAME C9 'devmath'
```

然后我们使用 MLn 的一些辅助功能来把这些数据填进去。第一个是 PUT,产生第一栏。第二与第三栏则利用 MLAV 来计算平均数。最后两栏则以 CALC 来进行简单的计算:

```
PUT 260 1 C5
MLAV C1 C3 C6
MLAV C1 C4 C7
CALC C8 = C3-C6
CALC C9 = C4-C7
```

为了 ANCOVA，我们需要进行虚拟编码来编定不同学校，为了不影响各栏的名称，我们输入下列文字：

```
DUMMY C1 C10-C19
```

到目前为止，我们已经有 19 个变量，已经足以用来分析本章的所有模型。首先我们先进行第 2.5 节的整体回归分析，我们必须告知计算机哪一个变量为解释变量，哪一个是反应变量：

```
EXPL   C3 C5
RESP   C4
IDEN 1 C2
IDEN 2 C1
```

进一步的建立模型，必须指出哪一个变量是固定效果，哪一个是随机效果。计算机预设所有的解释变量被设定为固定效果。CONS 是唯一带有方差成分的解释变量。

```
FPAR   C6-C19
SETV 1 C5
```

然后我们必须把批次（batch）模式设定打开，不然计算机就不会进行迭代估计，最后开始分析：

```
BATCH
START
```

2.11 摘　要

本章我们讨论了一些分析两层（个体层次与脉络层次）的分组数据的传统方

法。这些模型的数据分析都是在某一层中进行,不是个体层次就是脉络层次。在个体层次所进行的分析,会因为对变异部分的不同处理方式而有不同,导致不同的模型对于脉络效果的回归估计值的结果都有不同。从本章对于模型以及分析结果的讨论,我们可以得知我们需要较为一般化的模型。我们需要一个模型来分析数据所在的测量层次,在不考虑阶层结构或变量是如何聚合而来的情况下,来检验所有解释变量对于反应变量的影响,此种模型将在下一章讨论。

批 注

1. 在本书所使用的方程式中,我们使用底线来表示随机变量(Hemelrijk, 1966)。这不完全是一种正式的统计符号,但是我们认为此一标示有助于不同模型间的固定与随机系数的比较。

2. 此处有一个符号运用的问题。我们必须区分"数学成绩"这个名词的概念与在 NELS88 数据库的测量指针的标示。当我们在陈述概念的时候,我们直接以文字表述而不附加[ ]符号。如果是变量或标识,我们会在文字外部增加[ ]作为标示,例如[数学成绩]。

3. 我们以作为 $\stackrel{\Delta}{=}$"定义"的意思。

译者分析

1. **整体回归**:本范例利用 SPSS 软件进行回归分析语法如下(读者可以开启 School10. sav 文件,开启一个新的语法文件,复制下列语法后得到相同报表,也可以利用鼠标点选回归分析窗口来进行分析):

```
REGRESSION
  /DEPENDENT MathAch
  /METHOD = ENTER HomeWork /METHOD = REMOVE HomeWork .
```

对照表 2.3 的 SPSS 结果如下,其中模式 1 为带有[家庭作业]的模型,模式 2 为虚无模型。数据与表 2.3 的内容相同(数值如有差异应为计算过程取用小数位数与进退位的差异)。

模式摘要

模式	R	R 平方	调过后的 R 平方	估计的标准误
1	.497[a]	.247	.244	9.682
2	.000[b]	.000	.000	11.136

a. 预测变量:(常量),HomeWork time spent on math home

b. 预测变量:(常量)

系数[a]

模 式	未标准化系数		标准化系数	t	显著性
	B 之估计值	标准误	Beta 分配		
1 (常数)	44.074	.989		44.580	.000
HomeWork time spent on math homework	3.572	.388	.497	9.200	.000
2(常数)	51.300	.691		74.283	.000

a. 因变量:MathAch math score′

2. **加权聚合回归**:利用 SPSS 软件进行回归分析语法如下(数据文件为 School10_l2. sav 文件)。由于为加权回归,读者必须自行以数据库当中的 weight 变量进行加权,然后进行回归分析。

```
WEIGHT BY Weight
REGRESSION
  /DEPENDENT MathAch_mean
  /METHOD = ENTER HomeWork_mean /METHOD = REMOVE HomeWork_mean.
```

对照表 2.4 的 SPSS 结果如下,其中模式 1 为带有[家庭作业]的模型,模式 2 为虚无模型。数据与表 2.4 在标准误部分不同,在 R^2、斜率与截距部分则相同,显示作者对标准误的计算并非加权之后的估计值。读者可自行对照之。

模式摘要

模　式	R	R 平方	调过后的 R 平方	估计的标准误
1	.801[a]	.642	.640	4.414 38
2	.000[b]	.000	.000	7.360 16

a. 预测变量:(常量),HomeWork_mean HomeWork_mear

b. 预测变量:(常量)

系数[a]

模　式	未标准化系数		标准化系数	t	显著性
	B 之估计值	标准误	Beta 分配		
1 (常数)	37.109	.715		51.918	.000
HomeWork_mean HomeWork_mean	7.015	.326	.801	21.494	.000
2(常数)	51.300	.456		112.387	.000

a. 因变量:MathAch_mean MathAch_mean

3. **脉络模型**:虚无模型与脉络模型分析的数据文件为 School10_l1. sav(阶层一)与 School10_l2. sav(阶层二)。描述统计与 HLM6 分析报表列举如下。数值与表 2.5 相同。

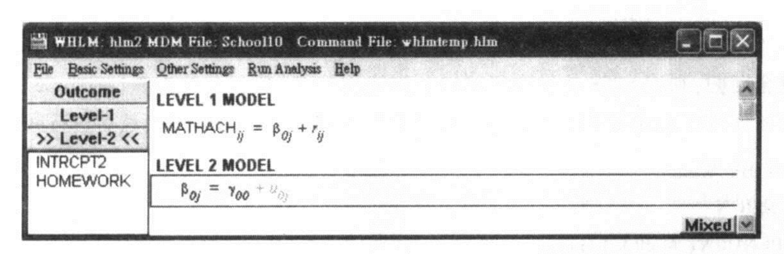
LEVEL-1　DESCRIPTIVE STATISTICS

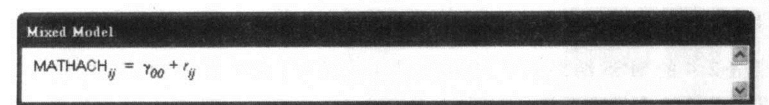
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
HOMEWORK	260	2.02	1.55	0.00	7.00
MATHACH	260	51.30	11.14	31.00	71.00

LEVEL-2　DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
HOMEWORK	10	1.76	0.70	0.86	3.30

虚无模型

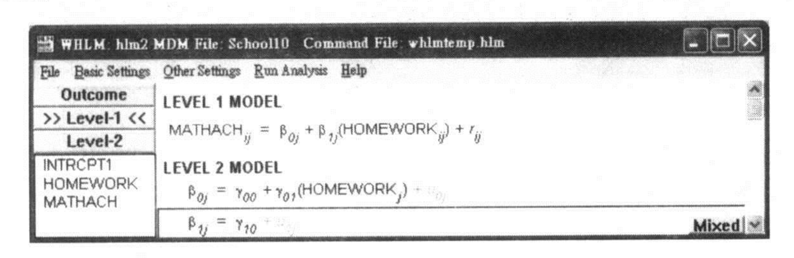




Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	51.300 000	0.690 603	74.283	259	0.000

脉络模型



Mixed Model

$MATHACH_{ij} = \gamma_{00} + \gamma_{01}*HOMEWORK_{j} + \gamma_{10}*HOMEWORK_{ij} + r_{ij}$

Final estimation of fixed effects:

脉络效果b_B　斜率b_W

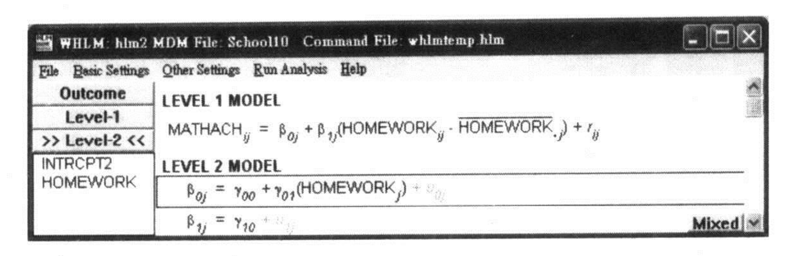
Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	41.431 209	1.708 614	24.248	257	0.000
HOMEWORK, G01	4.878 109	0.797 556	6.116	257	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	2.136 635	0.432 608	4.939	257	0.000

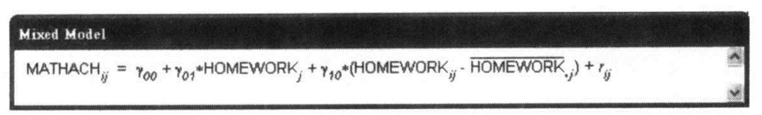
4. Cronbach **模型**:利用 HLM6.02 软件进行分析窗口如下,与前面脉络模型不同的是固体层次解释变量经过组平减处理。结果数值与表 2.6 相同。

Final estimation of fixed effects:

脉络效果b_B-b_W

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	37.108 633	1.467 442	25.288	257	0.000
HOMEWORK, G01	7.014 744	0.670 034	10.469	257	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	2.136 635	0.432 608	4.939	257	0.000





5. **协方差分析**：利用 SPSS 软件进行 ANCOVA 语法如下（数据文件为 School10_l1. sav 文件）。表 2.7 的虚无模型为不包含协方差的 ANOVA 模型，语法只需把第一行的 WITH HomeWork 删除即可。

```
UNIANOVA MathAch BY schoolid WITH HomeWork
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /PRINT = DESCRIPTIVE ETASQ OPOWER
  /CRITERIA = ALPHA(.05)
  /DESIGN = HomeWork schoolid .
```

对照表 2.7 的 SPSS 结果如下，R^2 数值相同。其余数据读者可自行对照之。

ANOVA

Tests of Between-Subjects Effects

Dependent Variable: MathAch math score′

Source	Type Ⅲ Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	14 030.536[b]	9	1 558.948	21.549	.000	.437
Intercept	545 077.651	1	545 077.651	7 534.498	.000	.968
Schoolid	14 030.536	9	1 558.948	21.549	.000	.437
Error	18 086.064	250	72.344			
Total	716 356.000	260				
Corrected Total	32 116.600	259				

a. Computed using alpha = .05

b. R Squared = .437 (Adjusted R Squared = .417)

R^2数值与表2.7同

ANCOVA

Tests of Between-Subjects Effects

Dependent Variable: MathAch math score′

Source	Type Ⅲ Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	16 034.203[b]	10	1 603.420	24.825	.000	.499
Intercept	178 018.604	1	178 018.604	2 756.220	.000	.917
HomeWork	2 003.667	1	2 003.667	31.022	.000	.111
Schoolid	8 100.396	9	900.044	13.935	.000	.335
Error	16 082.397	249	64.588			
Total	716 356.000	260				
Corrected Total	32 116.600	259				

a. Computed using alpha = .05

b. R Squared = .499 (Adjusted R Squared = .479)

R^2数值与表2.7同

变动与随机系数模型 3

3.1 绪 论

本章的主要目的在于说明变动系数模型(varing coefficient model)与较进阶的随机系数模型(random coefficient model; RC)之间的差异。本章使用与前章相同的数据库,也就是 NELS88 的 10 个学校的数据来说明随机系数模型的应用。在进行数据分析之前,首先利用四个假想学校的状况来说明这两种模型分析的异同,而两种模型的比较是在下列三个假设情境下进行:

1. 变动截距
2. 变动斜率
3. 变动截距与斜率

其中变动系数模型也被称为"斜率结果"取向("slope-as-outcomes" approach)。

利用四所假想学校的资料,我们可以发现两种模型具有相同的概念,但是 RC 模型是一个更具有统计学意味的变动系数模型。在概念的解释上,以"斜率结果"取向来说明会比较清楚,但是对于分析的结果,则以 RC 模型的结果较容易解释。此外,这两种模型具有相同的分析目的,但是 RC 模型更具有统计正确性、简约性以及便利性。

在介绍完这两种多层次模式的基本假设与公式之后,我们将把两种模型与第 2 章所讨论的传统回归模型与 ANCOVA 模型来一起比较。本章的最后,则以前一章的 10 个学校数据库来说明两种模型在分析结果上的差异。

3.2 分组回归

传统上，对于分组资料的分析，多利用不同形式的回归分析来处理，包括了ANCOVA。基本的线性方程式如下：

$$y_{ij} = a_j + b_j x_{ij} + \varepsilon_{ij} \tag{3.1}$$

公式3.1与第二章的公式2.2a相似。x是个体解释变量，y是反应变量，而a_j为截距，b_j为斜率。公式中各项系数使用复数型态，是因为每一个不同的组会有不同的系数，两个系数不会是单一数值。下标j加在a与b两个系数，表示第j组对于其中的各个个体i提供的共同的脉络背景。而ε_{ij}则为个体的误差项，为平均数为0、方差为σ^2的随机分布。在公式3.1中，仅有y_{ij}和ε_{ij}为随机变量，到了后面有关变动系数模型的讨论时，这两个变量也将保持有底线的状态（表现为随机变量）（注1）。

在NELS88的10个学校数据中，学生是个体层次的观察单位，学校则为不同的脉络（或分组），解释变量x为[家庭作业]，反应变量y为[数学成绩]。

3.3 变动系数

在传统的固定效果线性模型中，"斜率结果"取向可以被视为多阶层分析，也是通往进阶的多层次模型的第一步。在此取向中，由于有不同的学校（组），研究者利用一套带有个体解释变量与一个个体反应变量的线性模型来估计各分组下的参数。此时允许各脉络（亦即各校）拥有自己的微观模型（micro model）。图3.1到图3.3说明了四所学校的线性模型的三种不同情况。

基本上，四所学校具有类似的特性。在图3.1中，四所学校具有相同的斜率但是截距不同。图3.2中，四所学校具有相同的截距但是斜率不同。图3.3中，四所学校不仅斜率不同，截距也不同。

这三个图作为本章解释"斜率结果"分析取向的图示。在图3.1中，四所学校的回归方程式是平行的，平行意味着反应变量y（数学成绩）对解释变量x（家庭作业）作回归时，在各组中具有相等的斜率。但是由于这些直线的起点不同，表示各校数学成绩的平均水平有所差异。不相等的截距表示当个别学生的[家庭作业]状况被考虑进来后，有的学校数学成绩表现较好，有的较差。图3.1的状态即是一个ANCOVA模型，ANCOVA模型假设各组的x与y关联情形相同（回归线呈现平

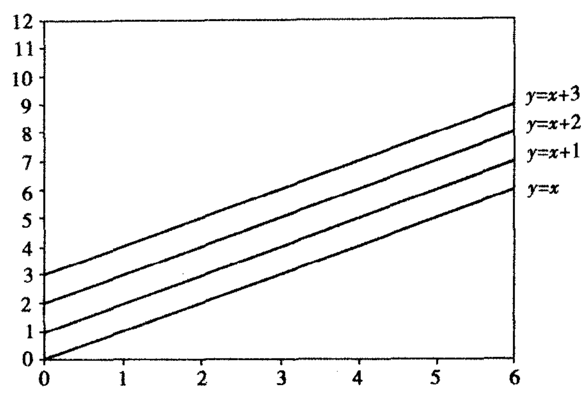


图 3.1 变动截距的四条回归线

行),来比较各组截距的差异。

图 3.2 的四所学校,回归直线有同一个起点,但是各有不同程度的斜率。也就是说,某些学校的 y 在 x 的回归较强,导致较大的斜率。越陡峭的直线,表示越强的数学成绩对家庭作业的回归效果。

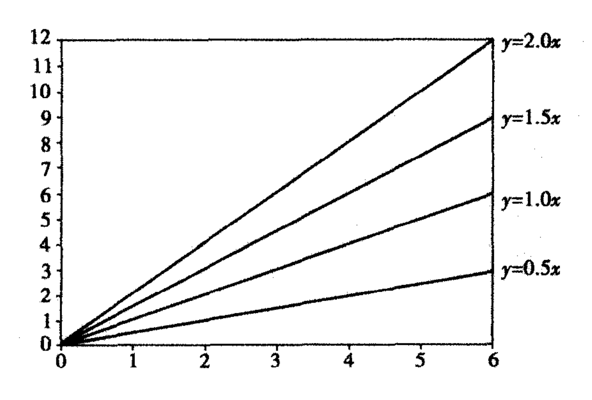


图 3.2 变动斜率的四条回归线

最后,图 3.3 中的四所学校状况可以说最符合实际的状况。四所学校的回归模型具有不同的斜率与截距,说明了"斜率结果"模型在分析上的重要性,四个学校都有独特的一套模型,比我们去强迫相信各校有相同的特性更符合实际。

这三个图形说明了分组数据的分析,每一个不同的组(学校)必须分别估计其斜率或截距,说明了"斜率结果"策略的第一步工作是去分别估计各组的方程式,然后才把估计得到的参数作为带有宏观层次解释变量的宏观层次回归的反应变量。

上述这种作法又被称为两阶段分析(two-step analysis),因为第一步先对个别的脉络进行微观层次的参数估计,然后作为第二步的反应变量,被宏观层次的解释变量所预测。在这两个步骤中,都以 OLS 法来进行参数估计(注 2)。

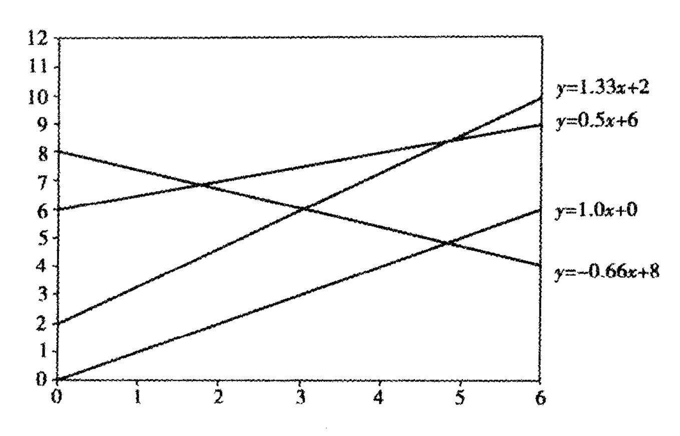


图 3.3 变动截距与斜率的四条回归线

两阶段分析的第二步方程式(宏观层次)如下:

$$a_j = c_0 + c_1 z_j \quad (3.2a)$$

$$b_j = d_0 + d_1 z_j \quad (3.2b)$$

其中 a_j 与 b_j 分别是截距与斜率两种回归系数,有关他们的应用与操作,可参考 Burstein 等人(1978)或 Tate(1985)的说明。

在两阶段方法中,每一个步骤的观察值个数不同。在第一步的微观层次中,各组内的观察值个数(各校学生数目)可能不同(如第2章的表2.1所列出的10个学校的各校人数)。在宏观层次的分析以斜率或截距为反应变量时,宏观层次的观察值个数则相同。在本范例中均为10,亦即10个学校产生10个截距 a_j 与10个斜率 b_j。宏观层次的截距(c_0 与 d_0)与斜率(c_1 与 d_1)如同公式3.2a与3.2b所示。这两个公式列为一组,表示都是以同一个解释变量 z 去解释截距与斜率的变化。在本章的范例中,此一解释变量为[公立]或整合得到的[家庭作业平均](注3)。

"斜率结果"取向是一种探讨传统数据分析上被忽略的关系或特性的有效策略。但是这种策略有一个实务上的缺点,是必须针对不同的脉络分别进行分析。分别估计所得到的数据或许可以用以描述个别脉络独特性,但是如果脉络复杂(学校数目众多),那么分析工作就非常繁杂而不够简效,更可能因而忽略了各校之间所具有的共通性。

为了解决上述的问题,最好的方法就是利用以下我们将要详细介绍的RC模型来进行分析。RC模型合并了前述"斜率结果"取向的优点与统计简效性,不但保留个别学校的特殊性,也兼顾了各校间的共通性。

3.4 随机系数模型

概念上，RC 模型是基于“斜率结果”模型的一种应用，所不同的是，虽然各脉络下的斜率、截距可能存在差异，RC 模型并不分别对个别脉络下的参数进行估计。图 3.1 到图 3.3 描绘四个学校的不同状态，每一个学校各自有其特色，因此可以建立自己的模型。为了进行比较，以下也举出对应的三个图形（图 3.4 到图3.6），来说明 RC 模型如何以单一模型来说明四个学校的状态。

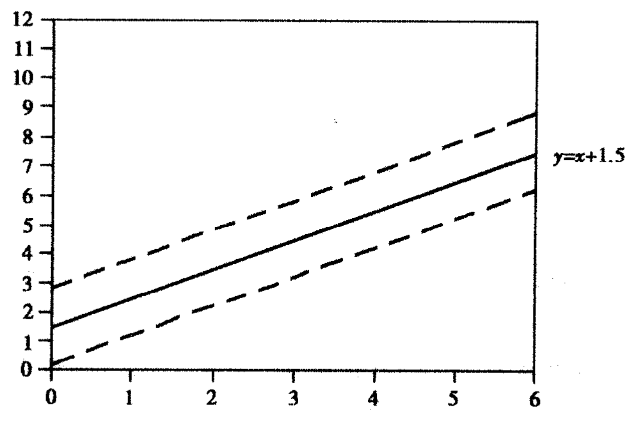


图 3.4 带有随机截距的随机系数模型

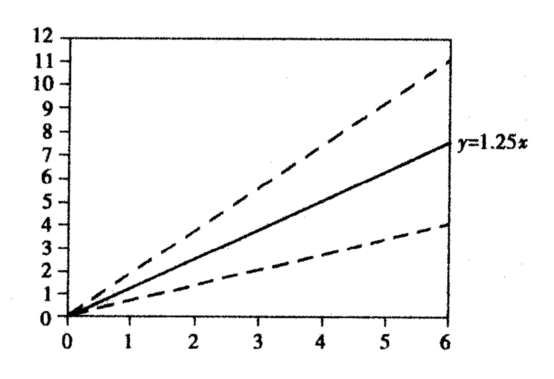


图 3.5 带有随机斜率的随机系数模型

在概念上，图 3.4 到图 3.6 等同于图 3.1 到图 3.3，所不同的是图 3.4 到图 3.6 中只有一条实线，在实线的两侧为两条虚线。这两条虚线说明的是四个学校从平均线离散的情形，相对应的就是前面图 3.1 到图 3.3 的“固定但变动的系数”的方差。在图 3.1 中，变动的是截距，图 3.2 与图 3.3 则涉及了斜率的变动，而图 3.4

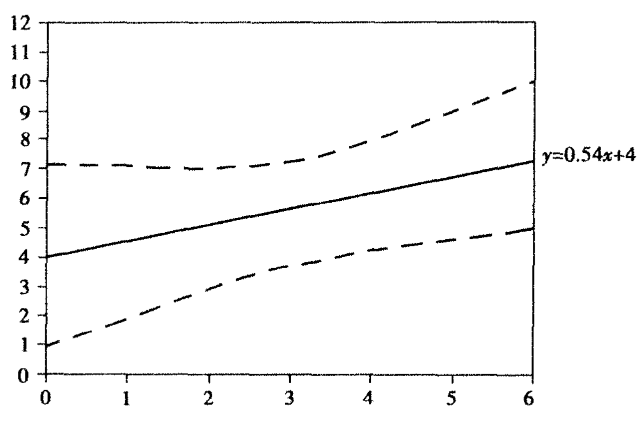


图 3.6 带有随机截距与斜率的随机系数模型

到图 3.6 也分别反应了这三种状况。

与图 3.1 相同,图 3.4 的截距变动但是斜率维持固定,但虚线表示了实线以外所可能存在的变异,而且是在不同的 x 之下,变异情形相同。同理,图 3.5 与图 3.2 均是斜率变动但截距不变的模型,而图 3.5 的虚线说明了不同的 x 之下,数据的变异情形不同。RC 模型所预期的就是这种结果,因为斜率的变动与个体解释变量 x 的数值有关,x 越高,总的回归线的散布区域就越大。

最后,图 3.6 与图 3.3 都是四所学校的斜率与截距双双变动的状态。在图 3.6 中,围绕平均线的散布就成了特殊的变动现象。虚线的趋势,合并了截距的方差与斜率方差,以及截距与斜率的协方差(covariance)。因为斜率的变动与 x 的数值大小有关,同样,协方差也会与 x 的数值大小有关。围绕在平均线周围的总方差,是这三个方差/协方差的加总,导致图 3.6 的散布在 x 变量最大值与最小值范围内,呈现不规则的状态。

如果平均线附近的散布点的方差(可由总方差来计算其数值)很大,我们即可以说该平均线无法等量地来反应各组的状况。由于该直线为平均的结果,我们从系数的离散情形或方差的大小可知,有不少学校高于或低于该直线。相对的,如果平均线附近的散布点的方差很小,该平均线就会等量接近各组的回归线。此时,利用单一层次回归方程式就可以反应各组的变量关系,反应得一样好。

值得注意的是,各学校有可能往上或往下变化,因为不仅斜率有别,截距不同,也可能两者同时存在差异,导致同一个学校可能在不同的 x 数值下,高于或低于平均线。

在 RC 模型,每一个系数都有方差,反映了各校的独特性。用统计的语言来说,每一个脉络(学校)的独特性由各校从平均线往外离散的方差来表示。在本章的后面,此一方差(或称为误差)可以用来计算后验平均数(posterior means)。后验

平均数是指每一个不同的脉络(学校)各自的斜率与截距的数值,非常接近"斜率结果"策略得到的结果。

图3.4到图3.6说明了RC模型的基本原则,也说明了RC模型与"斜率结果"模型的差异。以下,我们将把这些原则以数学方程式的形式来表述。图3.4到图3.6关于总的回归线的散布状况,说明了各脉络下个别估计得到的系数的差异性。由此可知,RC模型由两个部分所组成:平均数(固定部分)与方差(随机部分)。RC模型的随机部分由宏观层次的方差来描述,反应了统计模型从一个宏观模型会发生变动、变化的程度。此方差被用来指涉宏观层次的方差,之所以系数会不同,是因为组与组之间(各脉络间)的差异所造成,是高层而非微观层次的变异。

代表各校存在变动的随机模型的公式是从我们熟知的回归公式推演得到,如公式3.3(加底线者仍表示为随机变量):

$$\underline{y}_{ij} = \underline{a}_j + \underline{b}_j x_{ij} + \underline{\varepsilon}_{ij} \qquad (3.3)$$

下标 i 仍是指个体,j 仍是指组别,$\underline{y}_{ij}$ 是第 j 组的第 i 个人的反应变量的分数。x_{ij} 是同一个个体在个体层次解释变量数值,变量 $\underline{a}_j$ 是随机截距,变数 $\underline{b}_j$ 是随机斜率,ε_{ij} 是扰动项,ε_{ij} 被假设为期望值等于0(平均数为0),各扰动项 ε_{ij} 相互独立没有相关,ε_{ij} 的方差为 σ^2。

值得注意的是公式3.3的 $\underline{a}_j$ 与 $\underline{b}_j$ 所附加的底线是前面公式所不曾有的新特征,反应了这两个系数为随机系数。这也是公式3.3与公式2.2a("斜率结果"模型)的唯一差异。

到目前为止,我们所讨论的模型都有固定系数。在RC模型中,系数则既可以是固定的,也可以是随机的。而系数应为固定或随机的决定,在RC模型中,可以针对不同的系数分别考虑。

在RC模型中,系数被视为是带有方差的主效应,这一方差代表了所有脉络关于总水平或主效应的离差。更具体来说,所谓随机系数,是一个固定成分(fixed components)加上一个扰动项(disturbances)所组成。这些残差反映的是组别的变化,其特性类似于个体层次的 ε_{ij},被假设平均数为0的独立随机变量。

宏观层次的方程式,通过一个固定参数加上误差,来描述出随机斜率与随机截距的特性,如下所示:

$$\underline{a}_j = \gamma_{00} + \underline{u}_{0j} \qquad (3.4a)$$

$$\underline{b}_j = \gamma_{10} + \underline{u}_{1j} \qquad (3.4b)$$

公式3.4a与3.4b当中,宏观层次的误差 $\underline{u}_{0j}$ 与 $\underline{u}_{1j}$ 是指截距 γ_{00} 与斜率 γ_{10} 在脉络之间的变化。公式3.4a的平均截距为 γ_{00},$\underline{u}_{0j}$ 为每一个脉络截距与平均截距的

离差分数。同样,公式3.4b中,综合所有各脉络所估计得出的平均斜率为 γ_{10} , $\underline{u}_{1j}$ 为每一个脉络的斜率与平均斜率的离差分数。gamma 参数的下标,第一个数值是微观层次解释变量的编号,第二个数值表示宏观层次变量的编号。因此,对于 γ_{st} ,表示总体层次的第 t 个变量对于微观层次 x 变量回归系数的影响。0 表示截距,也就是说,不论在宏观或微观层次,该变量的所有数值均为 +1,例如 γ_{00} 表示宏观层次截距对于微观层次截距的影响。

值得注意的是,在公式3.4a与3.4b当中,系数 $\underline{a}_j$ 与 $\underline{b}_j$ 为固定效果 γ_{00} 与 γ_{10} 的函数,分别带着随机效果 $\underline{u}_{0j}$ 与 $\underline{u}_{1j}$ 。而 $\underline{u}_{0j}$ 的方差为 τ_{00} , $\underline{u}_{1j}$ 的方差为 τ_{11} , $\underline{u}_{0j}$ 与 $\underline{u}_{1j}$ 的协方差为 τ_{01} 。公式3.5列出带有一个随机截距与随机斜率的RC模型的方差成分:

$$T = \begin{matrix} \\ \underline{u}_{0j} \\ \underline{u}_{1j} \end{matrix} \begin{pmatrix} \underline{u}_{0j} & \underline{u}_{1j} \\ \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \tag{3.5}$$

公式3.5的T矩阵各元素为RC模型当中两个被估计参数 $\underline{u}_{0j}$ 与 $\underline{u}_{1j}$ 的方差与协方差。τ 表示各校从总平均离散的程度。

为了说明这些个别方程式不是真的被个别估计,而是模型中的一部分,我们可以将公式3.4a与3.4b带入公式3.3,得到下式:

$$\underline{y}_{ij} = (\gamma_{00} + \underline{u}_{0j}) + (\gamma_{10} + \underline{u}_{1j})x_{ij} + \varepsilon_{ij} \tag{3.6}$$

重组后得出:

$$\underline{y}_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + (\underline{u}_{0j} + \underline{u}_{1j}x_{ij} + \varepsilon_{ij}) \tag{3.7}$$

重组后的方程式显得更有组织,固定效果(gamma 参数)集合在方程式的前部,而微观层次的误差(ε_{ij})与两个宏观层次的误差($\underline{u}_{0j}$ 与 $\underline{u}_{1j}x_{ij}$)这三个随机效果则集合在后方的括号中。此一方程式除了括号中误差项之外,其他均与传统的回归方程式相同。在图3.6中也反映宏观层次斜率方差($\underline{u}_{1j}$ 的方差)与 x 数值的变化具有关联。公式3.7的 $\underline{u}_{1j}x_{ij}$ 即说明了此一关联性。

每一个脉络的独特变异(uniqueness)以宏观层次的误差变异($\underline{u}_s$)来表示,其意义为整体的模型解(回归方程式)的变异程度。对于各个脉络,并不会如前面的"斜率结果变量"策略一样,产生各自的模型解(回归方程式),在RC模型当中,只会产生一个整体的方程式。各校的模型解的变化,会围绕该单一方程式而变动,这条单一方程式的系数就是公式3.7当中的gamma系数,称为固定效果(fixed effect),宏观层次(学校间变化)的变异 $\underline{u}_{0j}$ 与 $\underline{u}_{1j}x_{ij}$ 则为随机效果(random effect)。如果这些方差显著不为0,我们即可宣称脉络的影响存在。

RC 模型的公式显示,该模型是一个介于完全限制(忽略脉络影响的传统一般回归)与完全没有限制(例如斜率结果变量策略,将脉络的影响完全纳入)的一种中间性的模型解。在斜率结果变量策略中,每一个脉络(学校)以一个独立的实体来处理,宛如它们之间没有任何共同相似之处;相比之下,传统一般回归模型则是将各脉络视为完全相同。从统计原理的角度来看,RC 模型也是介于两个极端之间,RC 模型所估计的固定参数比"斜率结果"策略为少,但是比一般回归模型多。下一节当中,我们将比较本章与第二章的模型的差别,并加以摘要整理。

3.5 线性模型的假设

表 3.1 摘述了回归和 ANOVA 两种传统线性模型与"斜率结果"和随机系数模型两种多层次模型的差异。表 3.1 中所列出的多数模型为固定效果线性模型,RC 模型是唯一的随机效果线性模型。在固定模型部分,截距可以为相等(公式 3.8a)或不相等(公式 3.8b):

$$a_1 = a_2 = \cdots = a_m \tag{3.8a}$$

$$a_1 \neq a_2 \neq \cdots \neq a_m \tag{3.8b}$$

表 3.1 传统线性模型与多层次模型的比较

模 型	截 距	斜 率
传统线性回归	相等	相等
协方差分析	不等	相等
"斜率结果"	不等	不等
随机系数	不等	相等或不等

公式 3.8a 为整体回归模型,该模型忽视了群体组别的差异,且在所有的脉络下,假设每一个个体都有相同的效果。在 ANCOVA 模型中,假设各脉络有不同的截距,如公式 3.8b 以及图 3.1 与图 3.4 所示。

线性模型对于斜率系数可以有不同的假设设定,斜率可以被设定为各组相等或不相等。在方差分析模型中,斜率则设定为相等,斜率的计算是用公式 3.9a 把各组斜率合并在一起而估计得出,随机与变动系数模型允许斜率为变动的数值,如公式 3.9b 与图 3.2 及 3.5 所示。

$$b_1 = b_2 = \cdots = b_m \tag{3.9a}$$

$$b_1 \neq b_2 \neq \cdots \neq b_m \tag{3.9b}$$

RC 模型与“斜率结果”模型允许研究者假设脉络内的系数与各脉络呈现特定函数关系的系统变化，不同的截距与不同的斜率可以设定在同一个模型中，如图 3.3 与图 3.6 所示。

比起上述两个多层次模型，ANCOVA 与回归模型是限定较多的模型。多层次模型由于放宽了一些限制让更多的参数得以自由估计，因此是较为一般化的模型。越一般化的模型虽然较限定模型有较多的自由弹性，但是简效性(parsimony)较差。在下一节，我们将讨论 RC 模型与“斜率结果”模型的特性，并以第二章的 NELS88 选取的 10 个学校数据库来进行实际范例的说明。分析的结果不仅进行两者的比较，也将与 ANCOVA 模型相比。

3.6 “斜率结果”分析

现在，我们来看看以“斜率结果”取向来进行阶层嵌套数据的分析范例。每一个脉络都有自己的线性模型，模型中具有一个或多个个体层次解释变量以及一个个体层次的反应变量。组内回归系数将在下一个阶段被拿来当作宏观层次回归分析的反应变量，各组的斜率与截距在模型中被汇整，成为后续宏观层次的反应变量。

我们以 10 个学校的实际资料来进行示范，因此可以得到 10 条不同的回归线，结果列于表 3.2 与图 3.7。图 3.7 绘出了 10 条回归线，实线表示 10 所学校，短虚线表示把 10 个学校合并在一起所估计得到的整体回归线。比较这 10 所学校不同的回归线与整体回归线，我们可以看出主要的不同在于斜率，截距的差异则较小。从表 3.2 所列出的相关系数(标示为 r 的数值)也可以看出[数学成绩]与[家庭作业]的相关系数的变化情形与斜率变化情形相同，显示出各校在[数学成绩]与[家庭作业]的关系上，具有实质的差异。虽然截距也有差异，但是就没有那么剧烈。事实上，当初我们在挑选样本时，就是选择这两个变项的关系有很大差异的 10 所学校，因此会得到如此的结果一点都不令人意外。基于 260 个观察值所计算出来的整体回归方程式的结果也列于表中。如果检阅图 3.7 的内容，可以发现短虚线与 10 所学校的回归线在斜率的差异远大于截距的差异。整体回归线的斜率呈现正向斜率，但是 10 所学校当中有 4 所学校的斜率是负的，显示出这些学校的斜率差异很大，以单一的整体回归线无法有效地反映这 10 所学校的状况。

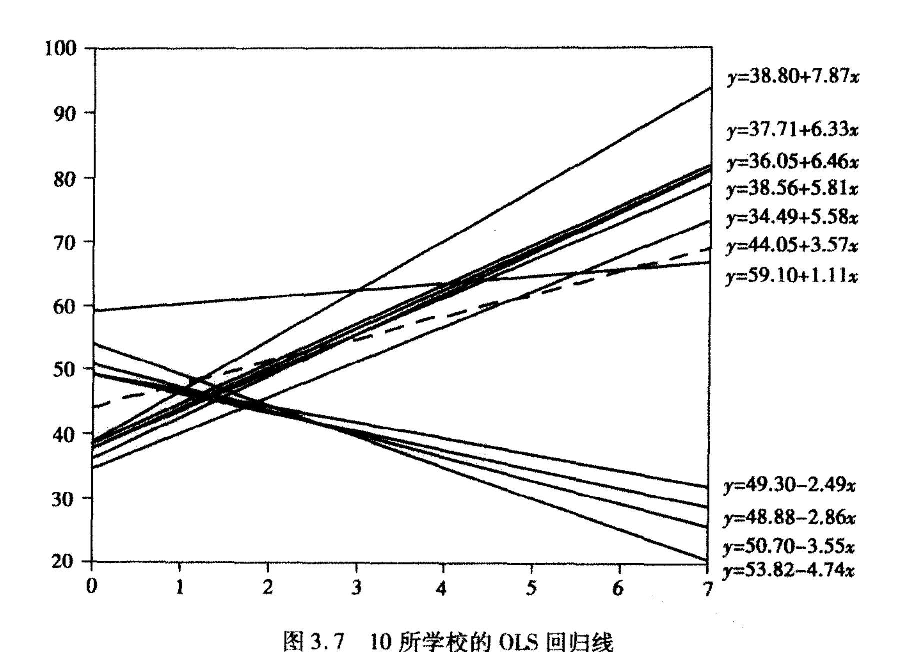


图 3.7 10 所学校的 OLS 回归线

表 3.2 10 个学校的数学成绩在家庭作业的 OLS 回归线[译者分析1]

学校	截距		斜率		r	N	Pu/Pr
	估计值	标准误	估计值	标准误			
1	50.70	2.24	-3.55	1.27	-0.52	23	1
2	48.88	3.56	-2.86	1.33	-0.45	20	1
3	38.80	2.94	7.87	1.37	0.77	24	1
4	34.49	1.76	5.58	0.80	0.84	22	1
5	53.82	2.55	-4.74	2.22	-0.43	22	1
6	49.30	1.51	-2.49	1.08	-0.48	20	1
7	59.10	1.42	1.11	0.38	0.34	67	0
8	36.05	3.46	6.46	1.46	0.71	21	1
9	38.56	3.19	5.81	1.99	0.56	21	1
10	37.71	2.36	6.33	1.11	0.80	20	1
全体	44.05	0.98	3.57	0.39	0.50	260	

在分析的第二个步骤，10 个学校的回归线的斜率与截距作为反应变量，被第二层的解释变量[公立]来加以预测，公立学校编码为 1，私立学校编码为 0，在公式 3.10a 与 3.10b 中，[公立]这个变量以 z 来表示。由于公立学校编码为 1，因此这个变量的系数反应的是公立学校的强度。

公式 3.10a 与 3.10b 列出两个不同的宏观层次模型,公式 3.10a 为截距部分,公式 3.10b 为斜率部分。

$$a_j = c_0 + c_1 z_j \quad (3.10a)$$

$$b_j = d_0 + d_1 z_j \quad (3.10b)$$

公式 3.10a 与 3.10b 两个公式产生宏观截距(c_0)与宏观斜率(d_0),[公立]主要效果的估计值(c_1),以及[公立]与[家庭作业]的跨层级交互作用(d_1)四个估计值(注 4)。

宏观回归的结果如下:

$$截距 \approx 59.10 - 15.95 \times [公立]$$

$$截距 \approx 1.11 + 0.94 \times [公立]$$

第一个宏观回归分析以截距为反应变量,其中[公立]变量的负系数表示公立学校的截距较低。第二个宏观回归分析以[家庭作业]的斜率为反应变量,[公立]变量的正系数表示公立学校的[家庭作业]斜率较私立学校的斜率更强(较陡峭)。在控制了家庭作业时数的影响之后,[公立]的两种学校类型的整体效果可以用公式 3.10a 来计算,私立学校的效果等于总体截距(59.10),公立学校的效果则较低,为 59.10 - 15.95 = 43.15;在[家庭作业]变量上的效果可以利用公式 3.10b 的第二个宏观层次回归公式来计算,这个公式带有一个微观斜率与[公立]的跨层级交互作用项。由公式 3.10b 可知,公立学校在[家庭作业]变项上有较强的效果,效果为 1.11 + 0.94 = 2.05,而私立学校的效果则为截距(1.11)。

在这个步骤中,[公立]变量有两种效果存在:一个发生在截距,称为整体效果(overall effect);另一个发生在斜率,称为跨层级交互作用(cross-level interaction effect)。

"斜率结果"取向的模型有一些缺点。第一,误差项结构的设定并不适当,导致参数估计的显著性(p-value)存有疑义。第二,第一阶段所获得的回归系数并没有相同的效力,有些系数的标准误较大,有些则较小。这些问题在第二阶段的分析过程并没有加以处理。在第二阶段的分析中,每一个系数的权重是相同的,但是"斜率结果"取向的概念仍然是有其需求。在下一节当中,我们将以相同的例子,说明"斜率结果"取向如何用来进行更具简效性的随机系数模型分析。

3.7 随机系数模型分析结果

对于随机系数模型的解释,我们仍然以同一个数据库来进行示范。[家庭作

业]为解释变量,[数学成绩]为反应变量,所使用的软件仍为 MLn(Rasbash et al.,1991)。分析的结果如下,系数下方的括号内数值为标准误:

$$[\text{数学成绩}] \approx 44.76 + 2.04 \times [\text{家庭作业}]$$
$$(2.59)\ (1.47)$$

若与固定回归模型的结果相比(表 3.2 的最后一行):

$$[\text{数学成绩}] \approx 44.05 + 3.57 \times [\text{家庭作业}]$$
$$(0.98)\ (0.39)$$

[家庭作业](x)与[数学成绩](y)之间仍具有正向的关系,但是却不显著,越多的做作业时间会有较高的数学成绩预测分数这一个说法不再有效。同一个系数在不同模型的数值差异虽然很小,但是由于标准误的差异很大,导致结论不同。在固定模型中,[家庭作业]的系数显著不等于 0,但是在随机系数模型中则不再显著。

宏观方程式考虑了回归系数的随机变异(见表 3.3)(注 5),前面所列出的单一方程式,可以改写成两个分离的方程式:

$$\text{截距}_j = 44.76 \times \text{宏观截距} + \text{误差}_j$$

$$\text{斜率}_j = 2.04 \times \text{宏观斜率} + \text{误差}_j$$

表 3.3　随机系数模型分析结果(随机部分)[译者分析2]

第二层		
参数	估计值	标准误
截距方差	60.89	29.89
[家庭作业]斜率方差	19.79	9.53
斜率与截距协方差	-27.91	15.32
第一层		
参数	估计值	标准误
方差(Variance)	42.89	3.92
偏差度(Deviance)	1 768.21	

在概念上,这两个宏观方程式可以与"斜率结果"取向的第二阶段相呼应。

一般来说,比起斜率,截距是较稳定的估计值。截距会有较大的方差,斜率方差则较小。然而,在我们人工挑选的数据库分析中得到的结果并非如此。在表3.3

当中，截距与斜率的方差大致具有相等的显著检验结果。虽然数值都不高，除以标准误后，一个是 $z=2.04$，另一个是 $z=2.08$，[家庭作业]（x_{ij}）与[数学成绩]（y_{ij}）的关系强度在各组间的变化与截距的变化（数学成绩的平均值）差不多。

3.7.1 增加一个宏观层次解释变量

以下，我们将讨论随机系数模型更进一步的分析策略。下面我们将增加一个第二层（宏观层次）解释变量[公立]进入模型之中。这个二分类别变项是用来“解释”截距与斜率系数的变化。通过增加一个学校层次解释变量 z，学校间在截距（普遍性）或在斜率（特殊性）部分的变异可能会消失。如果真的发生这种现象，我们称之为宏观层次变量“解释”了校间变异。在微观层次，我们使用相同的变量，以[家庭作业]预测[数学成绩]。

如同在“斜率结果”取向中，我们可以针对截距方差或斜率方差进行不同的模型设定。但是在“斜率结果”取向中，我们不会在同一个步骤中同时处理这两个方差的估计，但是在下面的例子中，我们将示范如何同时在模型中设定这两种宏观层次的方差。此时，我们不再像在“斜率结果”取向中，将宏观模型拆成两个模型，而是在同一个模型当中同时估计所有的参数。

我们的第一个工作是处理截距方差的解释。此时，宏观层次解释变量 z_j 被引入了截距方程式，但是并未引入斜率方程式。我们把公式 3.4a 修改成宏观层次变量[公立]（公式中的 z_j）与截距有关：

$$\underline{a}_j = \gamma_{00} + \gamma_{01} z_j + \underline{u}_{0j} \tag{3.11}$$

对于公式 3.4b 的斜率方程式则不进行任何变动：

$$\underline{b}_j = \gamma_{10} + \underline{u}_{1j} \tag{3.12}$$

在这个模型中，我们假设仅有截距的变动与宏观层次变量 z_j 具有函数关系再加上一个随机波动，这个随机波动代表宏观层次误差项 $\underline{u}_{0j}$（公式 3.11）。值得注意的是，我们所提出的模型为单一方程式模型，因此我们需将公式 3.11 与公式 3.12代入微观方程式（公式 3.3）之中，得到下列单一方程式：

$$\underline{y}_{ij} = \gamma_{00} + \gamma_{01} z_j + \underline{u}_{0j} + x_{ij}(\gamma_{10} + \underline{u}_{1j}) + \varepsilon_{ij} \tag{3.13}$$

加以重组后，得出：

$$\underline{y}_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + (\underline{u}_{0j} + x_{ij}\,\underline{u}_{1j} + \varepsilon_{ij}) \tag{3.14}$$

再一次的，公式 3.14 看起来就像一般固定效果回归方程式，但是带有复杂的误差项。固定部分的分析结果如下：

$$[\text{数学成绩}] \approx \underset{(2.67)}{57.98} + \underset{(1.52)}{1.93} \times [\text{家庭作业}] - \underset{(1.80)}{14.57} \times [\text{公立}]$$

随机部分的分析结果列于表 3.4。由表 3.4 可知，τ_{00}（截距方差）相当程度地被“解释”掉了。公私立之别具有显著负向效果就是证据。截距方差也发生了实质的改变，从 60.89 降为 40.20。这两个随机系数模型（无[公立]变量与有[公立]变量）的拟合度，可以利用两个模型的偏差度（deviance）的差异来评估（注 6）。判断的大略法则是两个模型的偏差度的差异必须是额外的估计参数数目的两倍以上。如果偏差度的差异达到显著水平，那么偏差度较小的模型可被视为是拟合度较佳的模型。应用这个原则到表 3.3 与表 3.4，我们得到偏差度的差为 18.73，自由度为 1（所增加的参数为 1）。很清楚，带有[公立]宏观层次变量的模型为较佳的模型，因为增加了这个变量，改善了模型与观察数据的拟合。

表 3.4　截距为宏观解释变量函数的随机系数模型分析结果[译者分析3]

第二层		
参数	估计值	标准误
截距方差	40.20	20.41
[家庭作业]斜率方差	21.58	10.30
斜率与截距协方差	-28.95	14.21
第一层		
参数	估计值	标准误
方差	42.78	3.90
偏差度	1 749.48	

我们的下一个任务是增加一个解释变量来解释学校间斜率的变异。从表 3.3 可知，斜率的方差达到统计的显著水平，因此值得进一步探讨是否宏观层次解释变量[公立]可以解释斜率的方差的随机性。在符号上，表示我们在宏观方程式（公式 3.4b）中增加一项 z_j：

$$\underline{b}_j = \gamma_{10} + \gamma_{11} z_j + \underline{u}_{1j} \tag{3.15}$$

我们还必须增加一项微观层次变量[家庭作业]与宏观层次变量[公立]的交互作用项，由于私立编码为 0，因此得到的结果是有关公立学校的状况。

将新的斜率宏观方程式与前面公式 3.11 的截距宏观方程式一起代入基本的微观方程式（公式 3.3），并加以整理后，得出：

$$\underline{y}_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}x_{ij}z_j + (\underline{u}_{0j} + x_{ij}\,\underline{u}_{1j} + \underline{\varepsilon}_{ij}) \tag{3.16}$$

模型 3.16 与模型 3.14 的差异仅在于一个新增的参数 γ_{11}，其余都保持不变。

固定部分的分析结果如下：

$$\begin{aligned}[\text{数学成绩}] \approx\ & \underset{(6.55)}{59.10} + \underset{(4.65)}{1.11} \times [\text{家庭作业}] - \underset{(6.92)}{15.83} \times [\text{公立}] \\ & + \underset{(4.92)}{0.92} \times [\text{作业} \times \text{公立}]\end{aligned}$$

随机部分则列于表 3.5 中。由固定部分的数据可知，跨层级交互作用项并没有达到统计显著水平。同样的结果也发生在两个模型拟合度的差异检定上，表3.4 与表 3.5 的偏差度差异为 0.04，自由度为 1，模型拟合度的改善程度并未具有统计的显著性。

表 3.5 截距与斜率为总体解释变量函数的随机系数模型结果[译者分析4]

第二层		
参数	估计值	标准误
截距方差	39.84	20.23
[家庭作业]斜率方差	21.37	10.20
斜率与截距协方差	-28.68	14.08
第一层		
参数	估计值	标准误
方差 Variance	42.78	3.90
偏差度 Deviance	1 749.44	

本节的主要结论是，带有随机斜率与随机截距的模型展现出较佳的模型拟合度，宏观层次的解释变量[公立]在微观层次截距中具有负向的效果。[公立]与[家庭作业]的跨层级交互作用对于模型并没有显著的改善效果。显著的随机效果仅部分能被学校特征[公立]所解释。

把随机系数分析与“斜率结果”模型所获得的结果拿来比较是非常有意思的。这两种模型所估计出来的截距与斜率参数都是相同的，分别是 59.10 与 1.11，宏观层次变量[公立]的参数估计也非常接近，分别是 -15.95 与 -15.83，交互作用项也是如此，分别是 0.94 与 0.92。主要的差别在于宏观层次变量与交互作用项的标准误的估计。

3.7.2 后验平均数

大多数统计软件对于随机系数模型所估计得到的结果是后验平均数(posterior means)。基本上,从"斜率结果"取向的第一阶段所得到的后验平均数可以用来进行组与组的比较。然而,随机系数模型的估计方法与"斜率结果"取向所估计得出的个别 OLS 回归分析结果非常不同。前面的示范中已经说明了 10 个学校分别进行斜率与截距估计的过程(请参考公式 3.11,3.14,3.15)。然而,随机系数模型估计方法则是经验贝叶斯最大似然法(EB/ML)(MLn 软件所使用的方法与 EB/ML 相同)。

在 EB/ML 估计过程中,考虑了整体回归模型的估计。而个别学校的模型则是不稳定的估计值,会围绕在整体模型四周呈现出缩动的状态。不稳定的估计值是由于组内样本数较少,或因为数据代表性不足,导致具有较大的标准误。当估计值不可靠时,会产生较大的缩动(shrinkage),当估计值可靠时,缩动则较小。由于缩动的存在,使得随机系数模型所估计得到的后验平均数与"斜率结果"模型估计所得到的结果(表 3.2)有所不同。如果比较表 3.2 与表 3.6(注 7),表 3.6 中的事后平均数是以整体模型估计结果加上某特定学校的 OLS 估计结果所得出,而不像"斜率结果"取向将每一个学校分开估计以获得后验平均数。标准误与每一个学校内的相关系数的定义不再那么单纯,因此表 3.6 中有关随机系数模型的估计结果就不再列举,而整体斜率与截距的估计值则再次出现在表中(最后一行)。

表 3.6 10 个学校随机系数模型的缩动 EB/ML 回归系数

学校	截距		斜率		样本
	估计值	标准误	估计值	标准误	
1	50.28	(-)	-3.14	(-)	23
2	48.77	(-)	-2.70	(-)	20
3	39.24	(-)	7.53	(-)	24
4	35.25	(-)	5.38	(-)	22
5	52.96	(-)	-3.75	(-)	22
6	48.62	(-)	-1.77	(-)	20
7	57.94	(-)	1.35	(-)	67
8	37.16	(-)	6.02	(-)	21
9	39.21	(-)	5.38	(-)	21
10	38.17	(-)	6.09	(-)	20
MLn	44.76	(2.59)	2.04	1.47	260

我们的范例所出现的缩动现象并没有很大,这是因为所有的参数的信度(reli-

ability)都很高。我们可以从每一个学校的模型中看到(表3.2),所有的OLS估计结果都具有统计显著性(因为我们挑选这10个学校时,就是挑那些具有显著关系的学校)。但是我们可以看到,缩动很大的学校,标准误都很大。例如,把两个表中的第5所学校与较稳定的第6所学校(或其他亦具有较可靠估计值的学校)的数据相比较,第5所学校估计出来的系数的标准误大了两到三倍之多。在图3.8中,后验平均数在代表整体回归模型的点虚线的上下变异,即说明了缩动的状况。点虚线代表的是γ的数值,或随机系数分析的固定效果。我们有时会用后验平均数来排学校的名次,但是因为缩动的存在,使得这些平均数的运用具有疑义。缩动意味着后验平均数很高(或很低)但样本少的学校会向总体平均数移动。最后很高与很低的学校甚至会有相同的排名。随机系数模型的一个弱点是不太适合应用到真实生活的组别(例如学校、组织)的决策研究,它的主要长处是用于社会科学的理论发展或检验数据特性。

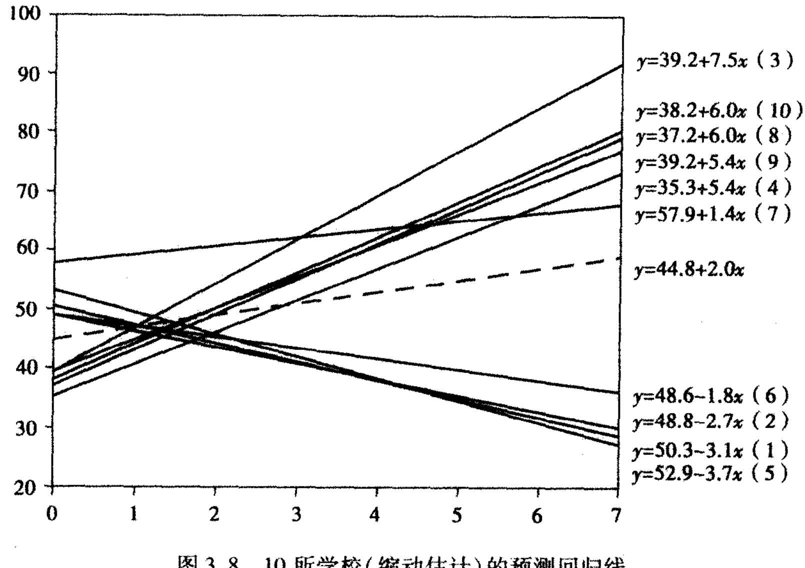


图3.8 10所学校(缩动估计)的预测回归线

3.8 替代模型:协方差分析

为了让我们的范例能够说明得更清楚,以下我们将套用ANCOVA技术在10所学校的数据库的分析中,看看会得到什么结果。在ANCOVA中,我们假设各组的斜率相等,但是很清楚,以我们的数据库的状况,这个假设是不成立的。

在ANCOVA中,每一个学校的截距估计值并不相同。以ANCOVA来分析我们

的数据,会得出一个[家庭作业]预测[数学成绩]的回归系数,称为总体组内回归系数(pooled within-regression coefficient),以我们的数据库所得到的结果如下:

$$[数学成绩] \approx 截距_j + 3.57 \times [家庭作业]$$

截距差异的 F 检定显示差异非常显著,$F_{(9,250)} = 13.84, p < 0.0001$,表示 10 所学校的某些或全部学校,在控制了做家庭作业的时数之后,数学成绩平均数具有显著差异。记得在具有组内相关的情况下,第一类型错误会有显著的扩大。经过我们的检验,组内相关为 0.30,非常高,以致于不能忽视(注 8)。在 ANCOVA 中,无法检验为什么这些学校的截距显著不同,我们所能做的是指出他们具有差别而已。是否这些学校的差异是因为公私立学校的差异所造成,无法用 ANCOVA 来检验,这类的结论已经超过了 ANCOVA 的范围。

3.9 参数的数目

随机系数模型在统计上的优点是可以明确地定义估计值的内容,此外,它们比“斜率结果”取向可以提供更为简效的模型。在我们的 10 所学校研究范例中,共有 30 个参数需要估计,每一个学校 3 个(截距、斜率、微观层次的方差)(参见第 3.2 节)。如果宏观层次解释变量与斜率、截距有关,模型另外要增加 6 个估计参数(见公式 3.2a),使得这个“斜率结果”模型的估计参数达到 36 个。对于同一个数据库,以随机系数模型(公式 3.15)来进行分析,则参数数目降到 8 个,即 4 个方差与 4 个固定系数。这 4 个方差是微观层次方差、两个宏观层次方差与它们的协方差;4 个固定系数为截距、[家庭作业]的微观斜率、[公立]的宏观斜率,以及跨层级交互作用[作业 × 公立]的斜率。如果把这两种模型加以简化,或许参数可以降到 8 个以下,例如把一些系数的方差设定为 0。

在我们的例子当中,可以把[家庭作业]变量的斜率的方差设为 0,让模型中只保留一个随机的截距,如此可以减少两个参数:斜率的方差,以及两个宏观误差(斜率的误差项与截距的误差项)的协方差。另一个作法则是保留[家庭作业]斜率的方差,因为在我们的范例中它具有统计的显著性。然后我们去掉[家庭作业]参数的固定部分。这个设定的决定,可以从两个斜率估计值(固定部分与随机部分)的 t 检验是否显著来判断。更好的方法是检验整个模型的整体拟合度,也就是计算放入变量与移除变量的两个模型的偏差度的差。由于两个模型的偏差度的差服从卡方分配,因此可以利用卡方检定来判断某一个模型是否优于另一个模型。

减少参数的方法还包括减少交互作用项的参数数目。在我们的范例中,公式 3.12 的模型比公式 3.15 的模型少了一个参数。在下面的章节当中,我们会回来讨

论这些参数增减的问题,以及针对全体的 1 003 所学校提出更合理的随机系数模型来进行分析。本章所使用的 10 个学校的例子只是全部样本的一个特殊组合(注 9)。

在以全体样本来进行分析的情况下,有一些设定上的差别必须说明。假设我们使用 ANCOVA 模型,亦即:

$$\underline{y}_{ij} = a_j + bx_{ij} + \underline{\varepsilon}_{ij} \tag{3.17}$$

所有的学校都必须假设具有相同的方差 σ^2 。在这样的模型中,只有一个变异成分,但是有一些其他的方差具有探讨的价值。第一是 m 个截距(a_j)的方差(可能要以各组样本数 n_j 来进行加权),此一方差的母体未知。第二个是 m 个截距($\hat{a}_j$) 的观察方差,得自于 ANCOVA 的最小平方估计值。我们可以暂时假设这个来自样本的方差可以反应母体的特性,但是它们之间的关系则非常复杂。

相同的,变动系数模型:

$$\underline{y}_{ij} = a_j + b_j x_{ij} + \underline{\varepsilon}_{ij} \tag{3.18}$$

模型中并没有斜率与截距的协方差参数,但是我们可以计算出斜率估计值与截距估计值的协方差,然后把它当作随机系数模型当中的协方差参数来处理。

因此,虽然 ANCOVA 与其他脉络模型并没有组间变异的方差成分,我们可以把具有相似功能的数据,利用 OLS 回归方法把他们估计出来后加以运用。

3.10 摘 要

本章中,我们比较了两种多层次模型:固定效果回归模型("斜率结果"模型)与更进阶的随机系数模型。前半段的篇幅我们从理论层次来讨论这两种模型的差异,我们挑选了四个假想学校来说明在研究上可能碰到的状况与应用上的涵义。本章的后半段,则是利用第 2 章的 10 所 NELS88 调查学校的数据来具体说明"斜率结果"取向与随机系数模型的分析过程与结果。文中我们详细讨论并比较了固定与随机模型的分析数据异同。

本章的主要结论是各模型间的固定效果估计值并没有很大的差别,但是实务上,这些模型要如何选择则有很大的差异。文中我们指出了随机系数模型的系数估计在缩动特性上所具有的优点与缺点。优点是会有较可靠的预测结果,缺点是对于样本数较少的学校、估计的结果较不切实际。这两个模型的差别在实务上的另一层涵义是随机系数模型较为简效,简效性使得随机系数模型的适用性较高,尤其是当一个研究要处理很多组别,或是每一组的样本数较少的时候。在本章的最

后部分,我们介绍了替代性的 ANCOVA 模型,对于截距视为固定效果的分析结果。很明显,ANCOVA 并不是一种分析脉络效果的严谨方法,ANCOVA 能做的是检验不同的脉络是否具有差异,但是却不知道差别的原因是什么。

批 注

1. 为了简化公式,本章仅以单一解释变量来进行示范,若要扩大到多元回归则很简单,只要多增加解释变量,即会得到更多的 b 系数。

2. 虽然“斜率结果”一词表面上看起来只是把斜率作为反应变量,但事实上截距也是反应变量,因此完整称呼应为“截距与斜率结果”。

3. 注意公式 3.2a 当中并不包括第二层误差项。虽然误差项并没有在模型中设定,并不表示截距与斜率为无误差或无变异。此处仅表示此模型并非完整定义模型。

4. 注意公式 3.10a 与 3.10b 当中,宏观方程式的误差项并不存在。在“斜率结果”模型中,宏观误差项并未设定在其中。稍后的随机系数模型中,将可看到宏观回归方程式中的误差项,此举也使得随机系数模型有较佳的统计优越性。

5. 注意,相对于“斜率结果”模型却少宏观误差项,在随机系数模型中的总体误差项则被适当地定义。

6. 嵌套模型之间所计算得出的偏差度差值服从卡方分配,自由度为两个模型参数估计数目的差值。

7. 注意表 3.6 当中的参数是基于单一随机系数模型(没有宏观解释变量)在单一一个步骤的估计结果,过程中并不会产生标准误的信息。

8. 我们再次看到第一章表 1.1 中所讨论到的组内相关对于第一类型错误膨胀的影响。表1.1 显示,当组内相关为 0.20 时,第一类型错误会从 0.05 膨胀到 0.46。

9. 注意,如果我们使用总数为 1 003 所学校的数据来进行分析,随机系数模型的参数数目不会改变,参数数目仍为 8。但是如果是“斜率结果”取向则不然,因为在第一个阶段必须对 1 003个个别模型进行估计,每一个模型有三个参数。

译者分析

1. **分组回归**:将 School10_l1. sav 就 10 个学校分别进行回归分析语法如下(结果与表 3.2 完全相同):

```
SORT CASES BY schoolid.
SPLIT FILE LAYERED BY schoolid.
REGRESSION
  /DEPENDENT MathAch
  /METHOD = ENTER HomeWork.
```

系数[a]

schoolid schoolid 模式			未标准化系数		标准化系数	t	显著性
			B 之估计值	标准误	Beta 分配		
7472	1	(常数)	50.684	2.211		22.920	0.000
		HomeWork time spent on math homework	-3.554	1.250	-0.527	-2.843	0.010
7829	1	(常数)	49.012	3.557		13.778	0.000
		HomeWork time spent on math homework	-2.920	1.330	-0.460	-2.195	0.042
7930	1	(常数)	38.750	2.943		13.166	0.000
		HomeWork time spent on math homework	7.909	1.375	0.775	5.753	0.000
24725	1	(常数)	34.394	1.796		19.150	0.000
		HomeWork time spent on math homework	5.593	0.818	0.837	6.835	0.000
25456	1	(常数)	53.939	2.523		21.377	0.000
		HomeWork time spent on math homework	-4.718	2.198	-0.433	-2.147	0.44
25642	1	(常数)	49.259	1.547		31.842	0.000
		HomeWork time spent on math homework	-2.486	1.108	-0.468	-2.244	0.038
62821	1	(常数)	59.210	1.431		41.366	0.000
		HomeWork time spent on math homework	1.095	0.385	0.332	2.842	0.006
68448	1	(常数)	36.055	3.464		10.407	0.000
		HomeWork time spent on math homework	6.496	1.462	0.714	4.445	0.000
68493	1	(常数)	38.520	3.189		12.079	0.000
		HomeWork time spent on math homework	5.860	1.989	0.560	2.947	0.008
72292	1	(常数)	37.714	2.367		15.935	0.000
		HomeWork time spent on math homework	6.335	1.116	0.801	5.678	0.000

a. 依变量:MathAch math score'

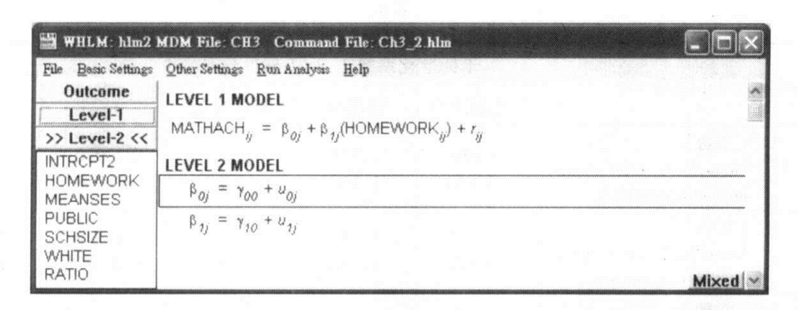
2. **随机系数模型(宏观层次无解释变量)**:数据库为 School10_l1. sav(微观层次),School10_l2. sav(宏观层次),以 HLM6 分析的操作窗口、描述统计与报表列举如下。

LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
HOMEWORK	260	2.02	1.55	0.00	7.00
MATHACH	260	51.30	11.14	31.00	71.00

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
HOMEWORK	10	1.76	0.70	0.86	3.30
PUBLIC	10	0.90	0.32	0.00	1.00



Tau

INTRCPT1,B0	69.304 26	-31.761 71
HOMEWORK,B1	-31.761 71	22.452 54

Tau (as correlations)

INTRCPT1,B0	1.000	-0.805
HOMEWORK,B1	-0.805	1.000

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	44.770 591	2.743 644	16.318	9	0.000

For HOMEWORK slope, B1					
INTRCPT2, G10	2.040 469	1.554 194	1.313	9	0.222

Final estimation of variance components:

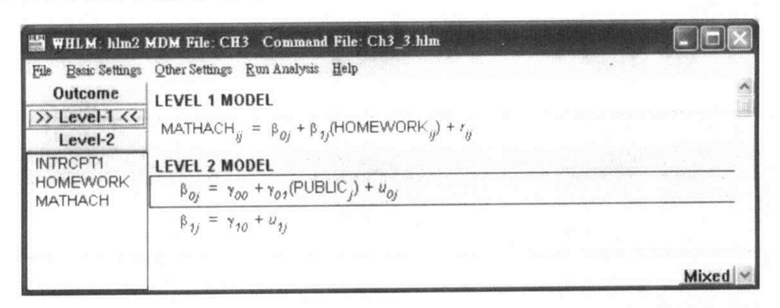
Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	8.324 92	69.304 26	9	153.181 27	0.000
HOMEWORK slope	U1	4.738 41	22.452 54	9	133.367 40	0.000
level-1,	R	6.562 85	43.071 00			

Statistics for current covariance components model

Deviance	= 1763.954 337
Number of estimated parameters	= 4

上述结果与表 3.3 的数值相近,斜率与截距数值接近。偏差度为 1 763.95,与表 3.3 的 1 768.21相近。

3. **随机系数模型(一个宏观解释变量** Public):在宏观层次加入一个解释变量 Public 对截距进行解释,并且带有抽样误差的随机效果。HLM6 分析的操作窗口与分析报表列举如下。



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	58.040 447	2.944 241	19.713	8	0.000
PUBLIC, G01	-14.660 674	2.109 003	-6.951	8	0.000
For HOMEWORK slope, B1					

INTRCPT2, G10	1.952 488	1.598 905	1.221	9	0.253

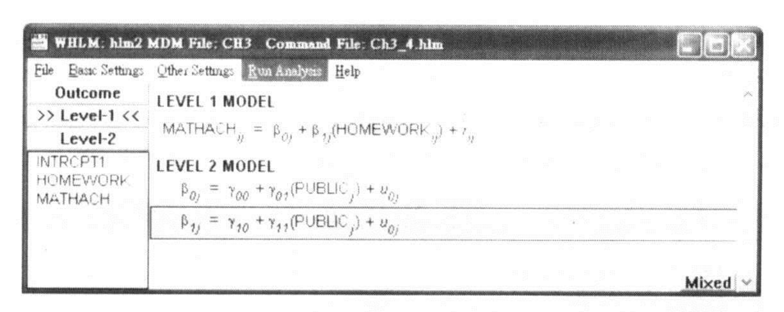
Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	6.77120	45.849 14	8	81.985 91	0.000
HOMEWORK slope, U1	4.895 07	23.961 76	9	133.963 27	0.000
level-1, R	6.554 40	42.960 22			

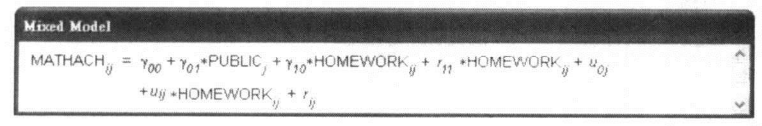
Statistics for current covariance components model

Deviance	= 1 742.128 440
Number of estimated parameters	= 4

上述结果与表 3.4 的数值相近，斜率与截距数值接近。偏差度为 1 742.13，与表 3.4 的 1 749.48相近。

4. **随机系数模型(一个宏观解释变量** Public)：在宏观层次加入一个解释变量 Public 对截距以及斜率均进行解释。以 HLM 操作窗口与分析报表列举如下。





Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	59.210 218	7.408 165	7.993	8	0.000
PUBLIC, G01	-15.965 966	7.830 068	-2.039	8	0.075
For HOMEWORK slope, B1					
INTRCPT2, G10	1.094 640	5.242 788	0.209	8	0.840
PUBLIC, G11	0.951 377	5.641 946	0.172	8	0.868

Final estimation of uariance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	7.200 74	**51.850 64**	8	81.547 49	0.000
HOMEWORK slope,	U1	5.221 88	**27.268 06**	8	129.578 45	0.000
level-1,	R	6.554 21	**42.957 62**			

Statistics for current covariance components model

Deviance	= **1 738.736 793**
Number of estimated parameters	= 4

上述结果与表 3.5 的数值相近，斜率与截距数值接近。偏差度为 1 738.74，与表 3.5 的 1 749.44相近。

范例分析 4

4.1 绪 论

本章我们将以 NELS88 数据为范例，讨论几种不同的分析方法。全章分为四个模型分析部分，分析的工具是 MLn 软件的二阶层分析。估计方法为未限制（即完全）最大似然估计法（FIML）（参见第 5.6 节）。

每一个部分的一开始都是从理论的观点出发，挑选为数不多的变量来进行示范。我们所挑选的几个基本解释变量（注 1）基本上都能够顺利产生成功的多层次分析结果。至于这些模型选择的理论意义与进一步的议题，例如变量的中心化议题等，则等到第 5 章再来讨论。在本章的范例中，我们将凸显多层次模型的一个通病，就是很容易让模型变得太大。大而复杂的模型或许看起来比较接近实际状况，但是要付出的代价是：不稳定性（instability）。不稳定意味着模型中一点点的变动，就会造成分析结果大大的不同，造成的理由很多，例如多元共线性就是其中之一。

我们以跨层级交互作用的分析为例，这种模型最适合用实际范例来说明它的运作特性。如果某个研究者有三个第一层解释变量与三个团体层次解释变量，此时，个体层次与团体层次解释变量的交互作用是十分具有研究价值的部分。个体层次的模型如下：

$$\underline{y}_{ij} = \underline{\beta}_{0j} + \underline{\beta}_{1j}x_{1ij} + \underline{\beta}_{2j}x_{2ij} + \underline{\beta}_{3j}x_{3ij} + \underline{\varepsilon}_{ij} \tag{4.1a}$$

对于截距的跨层级交互作用，为三个第二层解释变量 z_{1j}，z_{2j}，z_{3j} 所产生的三项俗称第二层变量的主要效果。方程式如下：

$$\underline{\beta}_{0j} = \gamma_{00} + \gamma_{01}z_{1j} + \gamma_{02}z_{2j} + \gamma_{03}z_{3j} + \underline{\delta}_{0j} \tag{4.1b}$$

而三个第一层预测效果（β_{1j}，β_{2j}，β_{3j}）与三个第二层解释变量 z_{1j}，z_{2j}，z_{3j} 的跨层级交互作用，总共会出现九个交互作用项：

$$\underline{\beta}_{1j} = \gamma_{10} + \gamma_{11}z_{1j} + \gamma_{12}z_{2j} + \gamma_{13}z_{3j} + \underline{\delta}_{1j} \tag{4.1c}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}z_{1j} + \gamma_{22}z_{2j} + \gamma_{23}z_{3j} + \delta_{2j} \quad (4.1d)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}z_{1j} + \gamma_{32}z_{2j} + \gamma_{33}z_{3j} + \delta_{3j} \quad (4.1e)$$

公式4.1c到4.1e三个方程式的九个交互作用项,分别与公式左边的三个回归系数有关。上述这些个别的方程式乍看之下不是在同一个步骤中所进行的估计,但是公式4.1b的主要效果参数与公式4.1c到4.1e的估计却是一次估计出来的。这些估计值多具有相关(见第5.7节),主要效果与交互作用有关,而交互作用彼此之间也有关系,导致经常到最后会发现,公式4.1c到4.1e的参数中,没有一个具有统计的显著性。如果是一个简单的模型,跨层级的交互作用系基于理论或基于我们对于样本数据的理解而设定,即可避免多元共线性的威胁,但是这样的模型或许不利于数据意义的解释与说明。

本章的各节示范了各种不同的分析方法,并说明我们有多种不同的选择。本章我们以RC模型为主,示范不同的操作方式以及结果的解释。有关模型建立的理论意义虽有交代,但是不会深入讨论,同时哪一个模型是最好的模型,无法从技术层次或模型拟合层次来说明,所谓最好的模型,是针对特定的目的而言。有关这种决定的选择的例子,请参考第5.2节的讨论。由于不同的模型得出不同的结果,模型的选择是非常重要的。甚至于一个研究必须是多层次模型而非回归模型,也可能是理论层次的决定。

4.1.1 数据描述

在以下的四部分的讨论中,[数学成绩]都将作为反应变量(亦即因变量)(注2),各部分一开始将使用少数几个学生或学校层次的解释变量来进行示范。一般在进行回归分析时,都会建议对于解释变量的数目进行限制,特别是当这些变量具有相关时。在多层次模型中,这个建议更加重要。在多层次模型中,参数数目的增加会非常快速,前面已经举了一个例子,若个体与团体层次各有三个变量,会产生九个交互作用项,合计所有的参数是解释变量数目的两倍以上。在各节当中,交互作用项与聚合变量都可能加入模型,其他可增加的参数包括随机斜率的方差与协方差,如此一来参数就会很多。就像其他所有的分析一样,多层次模型中重要变量的取舍也是取决于理论(theory),或是我们对于观察数据的理解(understanding)。“理论”与“理解”会引导分析的进行方式、变量的选择、参数的设定(随机或固定),或是否要估计交互作用的各种决定。

本章所进行的分析涉及七个变量,样本数据是从原来的NELS88的1 003所学校的21 580位学生中(注3)选取23个学校的519个学生,其中8个学校是私立学校、15个学校是公立学校。小一点的样本比较符合我们的需要,因为多层次模型的特性在较小规模的样本上比较容易得到符合真实世界的现象。

除了反应变量,我们选择了七个解释变量,学生层次(微观层次)的解释变量为:

- 社经地位(SES)[SES]
- 每周做作业的时数[家庭作业]
- 学生的种族背景:白人编码为1,非白人编码为0,因此这个变量我们称为[白人]
- 父母教育水平[父母教育]

学校层次(宏观层次)解释变量为:

- 学校种类[公立]:公立编码为1,私立编码为0,因此此一变量主要反应的是公立学校的状况
- 少数民族学生所占的比例[种族比例]
- 班级规模:为学生人数除以老师人数的[生师比]

在某些情况下会使用到聚合变量(微观解释变量的学校平均值),因此可能会有更多的学校层次解释变量。

虽然选择了这些变量,但是我们并不会把他们同时用在某一个模型中,因为这样一来,整个模型的参数数目会非常可观。如果所有可能的跨层级交互作用都被考虑进来,此一模型会有高达20个固定参数,分别是截距、[SES]、[家庭作业]、[白人]、[父母教育]、[公立]、[种族比例]、[生师比]、[SES]×[公立]、[SES]×[种族比例]、[SES]×[生师比]、[家庭作业]×[公立]、[家庭作业]×[种族比例]、[家庭作业]×[生师比]、[白人]×[公立]、[白人]×[种族比例]、[白人]×[生师比]、[父母教育]×[公立]、[父母教育]×[种族比例]、[父母教育]×[生师比]。毫无疑问,这些[个体变量]×[全局变量]的跨层级交互作用彼此都具有相关,也与主要效果有所关联。

如果每一个微观解释变量都设定为随机,模型的随机部分也将十分复杂。随机部分最多有10个参数:6个方差,4个协方差。方差分别是V(截距)、V([SES])、V([家庭作业])、V([白人]),协方差则是C(截距,[SES])、C(截距,[家庭作业])、C(截距,[白人])、C([SES],[家庭作业])、C([家庭作业],[白人])。

如此庞大的模型,很容易产生误导的结果。解释变量(包括跨层级交互作用)之间的相关使得参数估计值非常不可靠,使得模型当中的小小变化,或是换了一个样本之后,会得到很不一样的结果。

在本章当中,我们不会使用大模型,一般来说也不赞成读者使用。此外,我们建议在进行多层次分析之前,对于数据的特性进行必要的检查与分析,我们可以从

这个过程当中，理解数据的特性，进而引导我们去伪存真，发展一个只有几个解释变量的较小规模的模型。在本章当中，变量的选择是基于理论，而不是对于数据的理解。

学生层次解释变量的相关列于表4.1。结果显示即使是这些主要的变量之间也有颇高的相关，例如［平均 SES］与［SES］(0.65)，或是［SES］与［公立］(-0.35)。在表4.1中，以我们的样本所计算出来的相关系数列于上三角形，以原来的所有样本所计算出来的相关系数列于下三角形。两相比较之下，可以发现学校层级解释变量(［平均 SES］、［种族比例］、［生师比］)之间的相关在两组样本间的差异较大。比起1 003所学校的全体样本，我们所选择的23个学校有较大的标准误，是不令人惊讶的状况，更重要的是这23所学校是我们人为挑选出来的，而非随机取样得来，因此本章所得到的结果仅作为示范之用，不能推论到 NELS 88 的1 003所学校，甚至于全美国的学生母体。

表4.1 6个解释变量的相关

(上三角形为 N=519 与 23 所学校，下三角形为 N=21 580 与 1 003 所学校)

	［SES］	［家庭作业］	［白人］	［公立］	［平均 SES］	［种族比例］	［生师比］
［SES］		0.30	0.31	-0.55	0.70	-0.05	-0.30
［家庭作业］	0.29		0.09	-0.29	0.33	0.06	-0.16
［白人］	0.27	0.08		-0.10	0.24	-0.62	0.08
［公立］	-0.35	-0.12	-0.11		-0.78	-0.04	0.06
［平均 SES］	0.65	0.20	0.28	-0.54		-0.08	-0.42
［种族比例］	-0.23	-0.07	-0.59	0.18	-0.36		-0.22
［生师比］	-0.12	-0.06	-0.12	-0.12	-0.19	0.12	

在模型发展的过程中，我们以 MLn 的语法来说明模型的变化，避免使用公式。分析的结果，我们将以传统回归方程式的形式列出固定效果(固定参数估计值)，并在数据的下方列出标准误(括号内的数据)，然后以表格的形式，列出方差成分、组内相关(如果有的话)、偏差度等各项数值，同时将这些参数估计值依层级来表示(高层次放在上方，低层次放在下方)。

在分析过程中，我们从一个研究者的角度来探讨最佳模型是什么，因此对于变量的选择是基于理论知识，是否为最佳模型可以通过技术来检验(比较偏差度的差异)，替代的做法则是以某一个特殊理论观点所揭示的模型型态来进行分析，找出最佳模型。在本章中，我们示范了在第一个模型完成估计之后，还有其他替代的模型可以继续被提出来检验，在每一个步骤，我们会清楚地说明我们的考虑是什么。

4.1.2 本章四部分的组成

MLn 的第一个画面是空白的工作表画面，其中所键入的信息可以加以存盘再运用。以我们的范例而言，首先需以 NAME 指令命名七个解释变量的名称，在 NAME 指令之后，紧跟着字段编号（以 C 开头，依次为 C1，C2，…），接着就是以单引号写入变量名称。我们的数据库的第一个变量（C1）是学校辨识代码，第二个变量（C2）是学生辨识代码，我们分别以 school 与 student 作为两个层级辨识的依据，第二层是 school，第一层是 student，在 MLn 中是以 IDEN 指令来指定两个层级。指令撰写方式如下：

```
NAME C1 ‘school’ C2 ‘student’
IDEN 1 ‘student’
IDEN 2 ‘school’
```

在模型设定部分，我们需要指定一个反应变量与一个或多个解释变量。在 MLn 中，截距本身的性质是一个解释变量，在分析之前必须加以指定，由于截距为常数 1 的向量（注 4），因此模型设定是以常数的缩写（cons）来代表截距这个变量，并以 CODE 这个指令来指定 519 个观察值都有一个数值为 1 的向量区块，设定方法如下方的语法示范。在 CODE 指令之下，则将数值为 1 的这一栏命名为 cons，最后一个指令是设定常数与解释变量，常数在第一层被设定为随机（注 5）（以 SETV 1 指令来指定），在第二层也是随机（以 SETV 2 指令来指定）。本章所有模型的常数项均被设定为随机，第一层常数的方差将产生第一层误差项的方差，第二层方差则是截距方差。

```
CODE 1 519 1 C16
NAME C16 ‘cons’
EXPL ‘cons’
SETV 1 ‘cons’
SETV 2 ‘cons’
```

第一层方差随后可以被用来计算学校内的 R^2 。第二层的截距方差反应了学校层次的误差项的方差，随后被用来计算学校间的 R^2（注 6）。接着在所有的解释变量设定完成之后，包括一个或多个随机斜率，整个模型即可利用 MLn 软件进行分析。

本章的分析部分将会示范不同的多层次模型的应用方式。首先，在第一个部分，先分析虚无模型，接着是能够回答下列问题的模型："家庭作业时数是否是数学

成绩的良好的解释变量?",这个问题应改成"数学成绩与学生特征(例如家庭作业时数)的关系为何?"。在同一节中,我们将增加一个重要的解释变量:父母教育程度进入模型。值得注意的是,上述的研究问题完全与学校无关,从表面上来看,这并不是一个多层次研究问题,似乎我们并不需要进行多层次分析,但是,基于一些理由,这个部分的检验仍然有其必要。

首先,由于学生是嵌套在学校之内,亦即从学校中抽样得到,因此每一个学校内都可能存在着组内相关,多层次分析可以处理组内相关的问题。第二个理由是我们预期会有学校效果存在,学校效果不论是普遍影响各组或影响特定组,都可以在多层次模型当中来设定。对于各校截距的差异,可以利用 ANCOVA 模型来分析,但是至于为何学校间会有差异则需要利用多层次分析来回答。最后的研究结论可以利用变动系数模型(参见第三章)或多层次模型来分析得到。如果学校间的差异确实存在,不论是截距的差异或斜率的差异,下一个问题将是为什么会有差异? 这些差异是如何造成的? 这个问题 ANCOVA 模型并无法回答,因为学校的特性无法纳入模型来检验其影响效果,但是在多层次模型中,第二层解释变量甚至于跨层级交互作用就可以放入模型之中,用以解释学校之间的差异。

在第二部分,是将第一个部分的模型再加以扩充,加入了第二层解释变量与跨层级交互作用。微观层次解释变量[家庭作业]仍是[数学成绩]主要的微观层次解释变量。

在第三部分,则将环境(学校脉络)对于[数学成绩]的影响加入估计。一开始我们并没有放入第二层解释变量,但把学生的社经地位放入模型来预测[数学成绩],然后进一步把学校层次解释变量[种族比例]与[平均 SES]纳入模型。

在最后的第四部分,我们主要在探讨跨层级交互作用的影响,交互作用的分析需要在两个层次都有适当规模的观察值(这或许就是为什么我们子样本数据的分析结果没有统计的显著性)。于是我们以 NELS88 全部 1 003 所学校的数据来进行分析,最后得到了显著的跨层级交互作用。如果在较小规模的数据库,交互作用必须非常强,才会具有统计的显著性。我们以大型数据库来分析的结果,即发现大样本会得到较稳定的参数估计值,尤其是高层次的估计值。

4.2　第一部分

4.2.1　模型的标示

本章我们将以 MLn 的语法来描述被分析的模型。MLn 语法可以不需要利用任何符号或公式，而清楚描述某一个模型与下一个模型的差异。另一个优点是我们仅需利用一两个指令就可以将所需要的模型设定完成。根据我们使用多种软件的经验，MLn 可以让研究者有较高的操控空间。相对的，研究者必须熟悉 MLn 的相关选项与操作方法，就像我们在这一章所示范的（使用其他软件也一样），使用者必须非常清楚每一个指令的用途以及影响，才能够成功且正确地得到结果。为了让读者清楚每一个模型的变动，我们把 MLn 的指令标示在方框当中。

4.2.2　虚无模型（null model）

第一部分一开始是虚无模型（或零模型）的检验。一个虚无模型中，仅有一个反应变量，除了截距之外，没有任何解释变量。虚无模型被用来当做一个基准（baseline），作为后续进一步的模型当中“被解释”对应“未被解释”方差估计值的比较之用。同时，虚无模型也提供了对于反应变量[数学成绩]最初步的组内相关的估计值。在一个两层次的模型中，方差由两个部分所组成：个体层次方差与团体层次方差，或称为阶层一方差与阶层二方差。组内相关是截距方差（阶层二方差）除以总方差（阶层一与阶层二方差的和）而得。组内相关的另一个说法是组间变异所占的百分比。

组内相关的概念，是以仅带有一个随机截距的模型为基础。如果模型中带有任何解释变量，就无法计算出单纯、明确的组内相关系数。在一个带有随机斜率与随机截距的模型中，组间方差的数值是斜率与截距方差（与协方差）的合并。从最基本的随机系数模型（RC）的讨论中，我们可以得知反应变量 y 的方差与解释变量 x 的数值有关联。因此，在带有随机系数的模型中，带有不同 x 变量数值个体之间，他们的组内相关将有所不同，导致组内相关就不再具有相同的定义。一个虚无模型的组间与组内方差，可以作为估计“被解释变异量”（R^2）的标准（R^2 是传统回归的重要概念），亦即必须是仅带有随机截距的模型，不能带有随机斜率。传统的回归分析与二阶层的多层次分析差别就在于传统回归分析只有一个变异来源，因此只可以计算出一个 R^2。在一个两阶层的分析中，有两个可能的变异来源可以被解释变量来加以解释，因而可能得到两个 R^2（注 7）。阶层一的 R^2 是基于传统的误差变异概念所计算出来的数值，阶层二的 R^2（学校层次）则是一个新的概念。

MLn 对于虚无模型的设定需要一个指令,指出 cons 为解释变量,如下面方框所示。反应变量和第一层与第二层的截距 cons 的方差已经被先前的指令定义过了。

```
EXPL 'cons'
```

分析结果如下:

$$[\text{数学成绩}] \approx \underset{(1.13)}{50.76}$$

截距估计值得到一个数值为 50.76 的平均数,标准误为 1.13。方差与组内相关估计值则列于表 4.2。表 4.2 显示数学成绩的组内方差(81.24)比组间方差(24.85)大得多,在学校效能研究中,经常可以看到这种结果,反应了个别学生之间的差异远大于学校间的差异,或是说个体变异大于学校变异。组间(学校间)方差除以总方差(24.85 除以 106.09)就是组内相关,数值为 0.23。表 4.2 列出了偏差度(deviance),其数值等于负两倍的对数概似函数值(-2LL),可用来反应模型的拟合度,或是后续模型拟合度的改善程度的测量。

表 4.2 虚无模型(模型 0)结果[译者分析1]

第二层		
参数	估计值	标准误
阶层二方差	24.85	8.60
阶层一方差	81.24	5.16
组内相关 ICC	0.23	
偏差度 Deviance	3 800.78	

虚无模型的参数估计值将会在本章当中被重复提及。例如,当模型当中增加了解释变量与(或)随机斜率之后,虚无模型的第一层与第二层方差可以被用来当做方差减少程度(被解释程度)的参照点,而偏差度也有类似的功能。两个具有嵌套关系的模型,偏差度的差异服从卡方分配,两个模型参数的差异为自由度。当组间变异不再被清楚定义时(注 8),偏差度差异用于拟合度改善程度的评估显得特别重要。这就是为什么在分析结果的解释上,偏差度被认为是非常重要的数据,可以被用来评估模型的整体优劣。如果某一个模型的拟合度可以比另一个模型显著获得改善,一个经验法则是两个模型的偏差度的差值必须是两个模型估计参数数

目差异的两倍以上[1]。

4.2.3 [家庭作业]与[数学成绩]

本节是介绍在虚无模型当中增加一个解释变量成为模型1。模型的改变以下列方框来表示。新增的变量[家庭作业]是学生在家里做功课的时数,此一变量是以固定系数的方式进入模型(注9)。

```
EXPL 'homew'
```

固定效果的分析结果如下(括号内为标准误):

$$[\text{数学成绩}] \approx \underset{(1.14)}{46.35} + \underset{(0.28)}{2.40} \times [\text{家庭作业}]$$

把上述结果与虚无模型相比,我们可以发现截距数值产生改变,在[家庭作业]进入之前,截距为50.76。两个模型的方差成分(或随机部分)相比,也可以看到未解释变异量也减少了,随机效果列于表4.3。首先我们比较偏差度,虚无模型的偏差度为3 800.78,本模型(模型1)的偏差度为3 730.49,两者差异为70.29,模型1的估计参数比虚无模型多一个,亦即[家庭作业]的斜率,一个额外的参数表示模型1损失一个自由度,但偏差度的差值达到70.29,自由度为1,显示模型拟合的改善具有统计的显著性。

表4.3 模型1结果[译者分析2]

第二层		
参数	估计值	标准误
阶层二方差	20.23	4.70
阶层一方差	71.14	4.52
组内相关 ICC	0.22	
偏差度 Deviance	3 730.49	

在模型中增加[家庭作业]降低了第一层与第二层的方差,所降低的百分比,可分别用两个被解释变异量来计算(细节请参考第5.3节):个体的 R^2(以 R_W^2 表示)与学校的 R^2(以 R_B^2 表示)。组间变异从24.85降到20.23,差异为4.62,占学校层

1 如果以卡方分配的几率密度来判断,当 $\alpha=0.05$,自由度为1时,临界值为3.84,自由度为2时,临界值为5.99;当 $\alpha=0.01$,自由度为1时,临界值为6.63,自由度为2时,临界值为9.21。因此在正式报告研究数据时,建议以此卡方检验的原则来判断模型拟合改善程度的显著性。——译者注

次方差的19%;组内变异从81.24降到71.14,差异为10.10,降低幅度为12%。换言之,增加[家庭作业]解释变量得到的 $R_W^2=0.12$, $R_B^2=0.19$。

4.2.4 [家庭作业]的随机斜率

下一个模型(模型2)与上一个模型的差异也是在一个设定动作上:把[家庭作业]的斜率设定为随机,此外则没有任何新增变量。由于此一变动,我们预期各校的[家庭作业]效果有所不同。学校间的差异可能是学校规模的不同、管教方法的差异或是数学教材的差异所造成。若各校存在差异,则显示有些学校的学生数学成绩的高低与做作业的时间比较没有关系,有些学校学生的成绩好坏则与做作业时间比较有关。如果[家庭作业]在[数学成绩]的效果在某些学校真的比较弱,在某些学校比较强,我们预期会发现一个显著的[家庭作业]斜率的方差。在分析的操作上,只要增加一行语法,就完成此一设定,也就是把[家庭作业]的斜率以SETV 2设定在第二层。

```
SETV 2 'homew'
```

分析结果如下(括号内为标准误):

$$[\text{数学成绩}] \approx \underset{(1.72)}{46.32} + \underset{(0.91)}{1.99} \times [\text{家庭作业}]$$

把此一随机斜率模型(模型2)的结果与前面的模型1相比,我们可以发现两个系数的标准误都增加,截距则无变化。此一效果会在第5.5节中详加讨论。

在模型2当中,有两个额外的参数被估计:斜率的方差和斜率与截距的协方差,这些估计值与偏差度列于表4.4。分析的结果支持了我们先前的假设,亦即[家庭作业]与[数学成绩]的关系在各校之间具有差异,此一结论是基于显著的[家庭作业]斜率方差,方差为16.78,标准误为5.54,z 值为3.03。因为回归模型的参数估计值之间可能有相关,因此效果的整体意义必须就模型的整体拟合来检验,而不是就单一参数的检验来判断。如果模型的拟合改善程度具有统计的显著性,我们可以更安全地接受这个新增的参数具有意义。此一作法可以避免相关的参数的显著性检验问题。模型1与模型2的偏差度差异为3 730.49 - 3 639.04 = 91.45,自由度为2,新增的随机斜率可以有效地改善模式拟合度,2个自由度是指模型2新增的两个参数。研究结果显示带有随机的[家庭作业]斜率模型比固定斜率的模型有更理想的拟合。

表 4.4 模型 2 结果[译者分析3]

第二层		
参数	估计值	标准误
截距方差	59.28	20.00
斜率方差	16.78	5.54
斜率与截距协方差	-26.14	9.65
第一层		
参数	估计值	标准误
方差 Variance	53.30	3.46
偏差度 Deviance	3 639.04	

模型中加入随机斜率的设定，将增加两个参数，而非一个参数。由表 4.4 中，可以看到一个负的协方差，表示斜率与截距之间具有负相关。利用方差与协方差，可以计算出相关系数，数值为 -0.83（注 10），显示两者的关系非常高。为了避免斜率与截距的高相关带来不良的影响，可以把解释变量减去组平均数（组平减）获得中心化的解释变量，有关此作法及对于参数估计的影响可参阅第 5.2 节的讨论。

如果把模型 1 与模型 2 的结果相比较，我们会发现微观层次方差（组内方差）有一个预期外的降低，从 71.14 降到 53.30，由于模型 2 只新增一个第二层的斜率随机设定，在第一层并没有任何的变动，因此此一方差的降低并非我们预期的结果。有关第一层与第二层方差的复杂关系，我们在第 5.3 节中有较详细的讨论。同样的，这个问题的避免仍可通过[家庭作业]解释变量的中心化来达成（参考第 5.2 节），因为组平减程序会移除此一变项的组间差异，同时移除了与第二层参数之间的相关。但是，值得注意的是，组平减程序会改变模型的概念意义。[家庭作业]变量会获得一个特殊的组内变异与组间变异部分（[家庭作业]的平均值）。本节不详细讨论此点，而继续使用[家庭作业]的原始分数，把其他变量再加进来分析。

三个随机误差项具有相关，而反应变量 y 的方差与 x 变量的数值有关，因此方差不能直接加减，导致我们无法得到一个单纯的组间方差，也无法计算可解释变异量 R_B^2。另一方面，组内方差也会因为组间方差的变动而产生变化（见第 5.3 节），亦导致无法计算可解释变异量 R_W^2，也因此，无法计算组内相关。即在模型 2 这种最简单的随机系数模型中，R^2 与组内相关两个重要概念没有办法使用。

基于偏差度的数据，我们得到一个结论：增加[家庭作业]系数（斜率）的随机部分会使得模型有较佳的拟合。然而这样的一个较为复杂的模型，亦有其缺点，亦

即较不稳定。一般来说,较复杂的模型同时也是较不稳定的模型。在我们的例子中,不稳定性显现在[家庭作业]的参数估计上。我们可以比较前一个模型的[家庭作业]斜率的显著性检验,z 值为 2.40/0.28 = 8.6,表示是一个非常显著的系数;到了模型 2,则为 z = 1.99/0.91 = 2.19,仅微幅高于显著水平。

多层次模型的使用者必须理解一件事,即允许斜率在各脉络随机变动会导致系数本身的改变,在统计上,这是一种利害消长的现象,也就是我们获得了较佳的模型拟合,但是我们损失了个别参数的估计效率。基本上,这两种模型的选择,并不是基于模型拟合的好坏,而是理论与研究目的。如果学校的效果是研究的重点,那么随机斜率是很重要的一个设定。如果学校的效果不是重点,那么固定斜率可能是较佳的选择。本节的最后一个部分,我们示范固定斜率的这种比较传统的分析方法。

一开始我们会采用随机斜率,是因为我们关心学校的差异,[家庭作业]系数接近 0 这个现象,在多层次分析中并不会造成问题,更不是我们必须拿掉这个解释变量的理由。[家庭作业]这个变项仍然是对[数学成绩]的重要解释变项,但是是指它的随机效果而非固定效果。各校的斜率以 0 为中心存在着显著差异,显示[家庭作业]对于[数学成绩]的效果,主要是一种学校效果,而非微观层次的效果。

按照我们的逻辑,下一个步骤,我们应增加可以解释[家庭作业]变动斜率的学校特征变量(见第二部分的分析)。但是在进行此分析之前,我们先来看一下如果再增加一个学生层次解释变量[父母教育]的影响。

4.2.5 增加[父母教育]

在本节当中,所检验的理论观点是父母亲的协助会提升学生的数学成绩。这里所指的父母亲的协助,是假设父母的教育程度越高,对于学生越有帮助。因此,我们增加一个代表父母支持的解释变量[父母教育]到模型当中。在分析的设定上非常简单,我们只要在数据结构中增加[父母教育]这个变量,将它设定为解释变量(第八个变量,C8),并给予一个名称 pared 即可。

```
NAME C8 'pared'
EXPL 'pared'
```

此时这个模型有两个学生层次解释变量,但有一个随机斜率([家庭作业]的斜率),一个固定斜率([父母教育])。[父母教育]是微观层次解释变量,而非脉络变量(注 11)。新模型的固定效果结果如下:

$$[数学成绩] \approx \underset{(1.76)}{40.81} + \underset{(0.81)}{1.89} \times [家庭作业] + \underset{(0.29)}{1.85} \times [父母教育]$$

估计出来的数据出现了符合我们预期的结果,[父母教育]的斜率具有统计的显著性($z=6.38$),模型3的变异成分与偏差度的结果列于表4.5。表4.5中的偏差度数值降低了不少,表示模型3的拟合度颇佳。但是由于这个模型并非嵌套在前一个模型中,因此不能以偏差度的差值来进行模型拟合改善程度的卡方检验。同时,也如同模型2的限制,我们无法得到组内相关与解释变异量(R_B^2或R_W^2)的计算。为了示范,在下一节中,我们将以传统的回归分析来分析模型3的变项效果。

表4.5 模型3结果[译者分析4]

第二层		
参数	估计值	标准误
截距方差	45.20	15.69
斜率方差	13.08	4.41
斜率与截距协方差	-20.72	7.70
第一层		
参数	估计值	标准误
方差 Variance	50.70	3.30
偏差度 Deviance	3 602.35	

4.2.6 传统回归分析

传统的回归分析忽略数据的阶层结构特性。以我们的范例来说,学生是在各校之内来进行观察。若以传统回归分析来检验变项效果,我们先前所获得的组内相关系数$r=0.23$将被视而不见,而且前面的分析所发现的学校之间斜率与截距的显著差异的这种学校交互效果也会被忽视。一般而言,忽略组内相关将会导致回归系数标准误的低估,造成显著性检验高估与第一类型错误扩大的问题。有关理论层次的讨论,请参见第5.4节。以下我们将示范如何强迫所有的学校以单一一条回归线,来进行[家庭作业]与[父母教育]对于[数学成绩]的预测,这种做法将损失大量的重要信息,导致不良的模型拟合。

在MLn软件中,传统回归分析的设定是将截距与斜率的方差设定为0,以CLRV指令消除截距cons与[家庭作业]斜率在第二层的方差:

```
CLRV 2 'cons' 'homew'
```

上述指令把模型3随机系数模型(RC)改成传统的回归模型,我们可以藉此了解两个模型的差别。RC模型的回归系数变异成分被上述指令消除,形成下列大家所熟悉的固定系数回归模型。

$$[\text{数学成绩}] \approx \underset{(0.99)}{37.24} + \underset{(0.27)}{2.34} \times [\text{家庭作业}] + \underset{(0.28)}{3.00} \times [\text{父母教育}]$$

此一模型(模型4)的固定效果数据与前面RC模型的结果有所不同。传统回归模型的[家庭作业]系数数值,接近无随机斜率的多层模型的结果(表4.3),甚至标准误也十分接近。[父母教育]系数的强度在本模型则较高。三个回归系数的标准误中,有两个如预期的比RC模型来得小,传统回归模型的拟合也如预期般的较低。模型4的偏差度(见表4.6)显著高于RC模型(模型3),差值为115.94,自由度为2。表4.6中,由于传统回归模型仅有一个误差源,因此我们只会看到一个个体误差方差。

表4.6 模型4结果[译者分析5]

参数	估计值	标准误
阶层一方差	75.68	4.70
偏差度 Deviance	3 718.29	

固定性的回归分析,可以回答学生层次的研究问题,也可以回答学校层次的研究问题。例如我们把变量聚合成学校层次的数值,即可进行学校层次的分析。但是,固定回归分析一次能处理一个层次。多层次模型则可同时回答两个层次的问题,例如"各校的数学成绩与家庭作业的关系是否相等或不相等?如果不相等,那又是为什么?"。在我们的范例中,在模型中增加[家庭作业]随机斜率提高了模型的拟合度,所得到的结论是一个有关学校效果的假设检定。现在,我们即可以加入学校层次解释变量,来探讨为何[家庭作业]的系数会在学校间存在显著差异。

4.3 第二部分

4.3.1 简 介

在这个部分中,我们将介绍学校层次解释变量,我们挑选了两个重要的学校特征:学校的大小[学校规模]和学校类型[公立]。先前的分析发现,[家庭作业]的截距与斜率的显著变异可能可以被学校间的差异来解释,但是是哪些学校特性可以为这些变异负责还未知。我们所能做的,是去检验我们数据库当中的学校层次

变量是否与这些我们观察到的学校间变异有关。如果我们加入了这些变量进入模型后,斜率或截距的校间差异消失了,即表示我们找到了一些可以解释校间差异的原因。

图4.1当中,我们以大家熟悉的路径模型来说明多层次分析当中的学校特征可以影响低阶层关系的不同途径。图4.1当中,一些从路径模型发展出来的特殊概念,恰可以用来说明多层次模型的特殊功能,而利用路径图的方式,可以让读者从视觉层次而非公式,来理解我们的数据的各种关系的意义。如果读者希望把这个图形的内容与方程式加以连接,我们在连接相关变量的路径旁边标出了相关的符号,读者可以自行建立相对应的方程式。

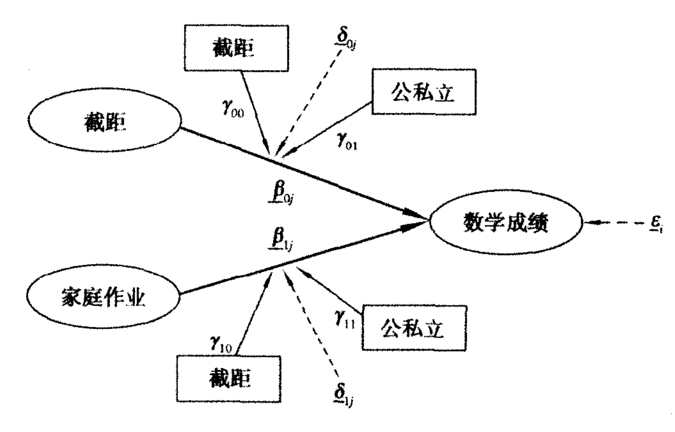


图4.1 随机系数回归模型

在图4.1中,我们呈现了一个带有一个学生层次解释变量([家庭作业])与一个学校层次解释变量([公立])的多层次模型。路径模型中有方框也有圆圈,以及一些看来不寻常的路径。圆圈代表学生层次的变量,方框代表学校层次的变量,从第一层次解释变量指向反应变量[数学成绩]的箭头表示影响的方向,其他箭头表示第一层的影响被第二层的变量所干扰。

第二层的特征总是与第一层的特征产生交互影响,我们以指向路径而非指向变量的箭头来表示,这些不寻常的箭头即表示多层次分析所具有的独特的跨层级交互作用,我们可以利用路径分析的直接与间接效果的概念来理解这些箭头的关系。

在多层次分析中,第二层变量是模型中的一部分,这些变量会与第一层变量产生交互作用,包括对于从截距指向反应变量的路径的影响,以及对于从第一层解释变量指向反应变量的路径的影响。在我们的例子中,学校层次变量[公立]有两个跨层级交互作用,一个是与截距指向[数学成绩]的路径所发生的交互作用,另一个是与[家庭作业]指向[数学成绩]的路径所发生的交互作用。

在路径模型当中的γ系数表示模型中的参数,可以用来呼应前面章节所提到

的参数。注意第二层变量在第二层也有常数(截距)项,而 γ_{00} 与 γ_{01} 两个系数取代了从截距指向反应变量的 β_{0j} 系数,而 γ_{10} 与 γ_{11} 两个系数取代了从[家庭作业]指向反应变量的 β_{1j} 系数。而虚线箭头表示误差项,可从方程式中计算得知。

因为[学校规模]与[公立]是学校层级变量,在模型中我们不允许这两个变量具有随机系数(注 12),本章的所有讨论都只允许学生层次的变量可具有随机系数,也就是允许这些第一层变量的系数可在学校间存在变动。

4.3.2 带有[学校规模]的模型

有关学校效能的文献指出,较小型的学校在提升学生学习成绩上会有较佳的表现。基于这个假设,我们在模型中增加了一个学校层次解释变量[学校规模]。我们另外开启一个新的 MLn 语法,重新建立模型。我们定义反应变量(RESP 指令)为[数学成绩],定义两个解释变量(EXPL 指令)[家庭作业]与[学校规模],以截距(cons)来定义方差(SETV 指令)分别出现在第一层与第二层,另外加上一个第二层方差,即第一层解释变量[家庭作业]斜率的方差。整组指令如下:

```
RESP 'math'
EXPL 'homew' 'schsize'
SETV 1 'cons'
SETV 2 'homew'
```

分析结果如下:

$$[\text{数学成绩}] \approx \underset{(2.62)}{44.95} + \underset{(0.91)}{1.99} \times [\text{家庭作业}] + \underset{(0.62)}{0.43} \times [\text{学校规模}]$$

此一模型(模型 5)具有两个解释变量,一个学生层次,一个学校层次。分析结果再一次显示[家庭作业]具有接近显著水平边缘的效果,如同我们先前所讨论的,[家庭作业]的显著的固定效果在加入了[家庭作业]系数的随机效果后就跌到显著水平的边缘,在这个模型中也是如此。新增变量[学校规模]并没有统计的显著性。表 4.7 显示了模型的方差成分与标准误,模型的偏差度报告于表格的下方。研究的结论指出,[学校规模]并没有显著的效果,与表 4.4 的模型相比,模式拟合也没有提高。本节范例仅增加一个参数,亦即[学校规模]的系数,表 4.7 的偏差度为 3 638.61,而没有[学校规模]的表 4.4(模型 2)的偏差度为 3 639.04,差值为0.43,改善幅度非常微小,因此我们的结论是[学校规模]对于[数学成绩]没有影响。

表 4.7 模型 5 结果(从模型 2 增加[学校规模]变量)[译者分析6]

第二层		
参数	估计值	标准误
截距方差	62.22	20.90
斜率方差	16.84	5.56
斜率与截距协方差	-27.27	9.94
第一层		
参数	估计值	标准误
方差 Variance	53.30	3.47
偏差度 Deviance	3 638.61	

4.3.3 以[公立]代替[学校规模]

在本节当中,我们把[学校规模]去除,改以[公立]变量放入模型中。在原始的数据库中,有一个变量[学校类型]为具有五个选项的类别变量(参见附录),其中一个选项是公立学校,其他四个选项为私立学校,我们的目的是比较公立与私立学校的差别,因此将该变量重新转换成一个具有两个水平的虚拟变量,公立学校编码为 1,私立学校编码为 0,命名为[公立],这个变项的效果反映了公立学校对比其他私立学校的结果。在本模型中(模型 6),研究问题是公立学校(相对于私立学校)是否能够解释前面模型所观察到的显著的截距与斜率方差。文献指出,公立学校的学生平均而言,学习表现会低于私立学校的学生,因此我们预期[公立]变量会得到一个负向的回归系数。

以下的指令显示我们消除了[学校规模],但是新增了[公立]变量:

```
EXPL 'schsize''public'
```

固定效果部分的结果如下:

$$[\text{数学成绩}] \approx \underset{(2.11)}{49.06} + \underset{(0.90)}{1.98} \times [\text{家庭作业}] - \underset{(1.90)}{4.08} \times [\text{公立}]$$

分析结果显示,[公立]变量确实如假设般地出现一个负向的回归系数,而且具有统计的显著性,$z=2.15$。随后我们会以范例说明这个效果并不是十分稳定,因为模型整体拟合并没有因为新增[公立]这个变量而有多少明显的改善。我们先把[公立]当作是一个有意义的变量,负向符号代表私立学校的学生,他们平均来说数学成绩较高。所谓平均来说,是指私立学校的方程式截距较高,私立学校的截距可

由截距减去[公立]的系数而得到:49.06 - 4.08 = 44.98。私立学校截距比公立学校截距高了 4.08,亦即平均来说,当大家都有相同的做作业时数的情况下,私立学校学生的数学成绩比公立学校学生高 4.08 分。上述说法是就数值本身的大小来解释,是一个大胆的说法,因为公私立学校的 4.08 分差异可能不是一个具有显著性的差异量。表 4.8 与表 4.4 的偏差度相比,发现两个模型的偏差度差距为 3 639.04 - 3 634.84 = 4.20,这么小的拟合度差异量(即使显著),提醒我们对于[公立]变量的效果的解释要特别小心。另一方面,由于我们的数据库并不大,只有 23 个学校,在组别较少的情况下,对于学校层次效果进行检验的统计检定力会偏低,除非效果非常强(详细讨论请参见第 5.4 节)。

表 4.8 模型 6 结果[译者分析7]

第二层		
参数	估计值	标准误
截距方差	56.24	19.10
斜率方差	16.37	5.42
斜率与截距协方差	-25.99	9.44
第一层		
参数	估计值	标准误
方差 Variance	53.34	3.47
偏差度 Deviance	3 634.84	

4.3.4 增加[公立]的跨层级交互作用

在文献中,我们可以发现公私立学校有另一种效果,亦即私立学校比公立学校更重视平等主义(egalitarian)。Raudenbush & Bryk (1986)的研究指出,社经地位(SES)与学生的少数民族背景在私立学校比较没有预测力,在公立学校则较强。换言之,公立学校的平等主义较为淡薄,较倾向于精英主义,此种型态的效果,在多层次分析中称之为跨层级交互作用。

在进行分析之前,我们先对回归模型的交互作用做个简单的回顾,因为在多层次模型中交互作用的效果更为特别。一般而言,在回归模型中加入交互作用意味着不稳定性的增加(Aiken & West, 1991)。交互作用项,例如公私立与社经地位的交互作用项,与原来的两个变量具有相关。解释变量间的相关形成多元共线性问题,是造成模型不稳定的非常重要的因素之一,有关此议题的细节讨论请参见第 5.7 节。

在我们的数据中,[作业 × 公立]交互作用项与[家庭作业]的相关为 0.30,与

[公立]的相关为0.65,见表4.9。对于[公立]与[家庭作业]两变量的高相关并不是本节所重视的问题,重点在于跨层级交互作用项的相关。表4.9所列出的相关系数是[家庭作业]与三个第二层解释变量的跨层级交互作用项的相关系数值。三个第二层解释变量为[公立]、[平均SES]、[生师比]。所有的系数皆指出同一个结论,亦即跨层级交互作用不仅与第一层解释变量有明显相关,也与第二层解释变量有高相关。表4.9当中括号内的相关系数是以NELS88的全部样本所计算得出,在所有的数据中,都至少有一些相关很高。如果我们说交互作用项会导致参数估计的不稳定,确实是有其根据。此时,如果对第一层变量进行中心化处理,可能会降低其中一些相关系数的数值,相关的细节将在第5.2节中讨论。

表4.9　跨层级相关矩阵(括号内数值为以NELS88全部数据所求得)

	[家庭作业]×[公立]	[家庭作业]×[平均SES]	[家庭作业]×[生师比]
[公立]	0.83(0.81)		
[平均SES]		0.83(0.81)	
[生师比]			0.27(0.31)
[家庭作业]	0.30(0.70)	0.43(0.21)	0.86(0.89)

现在,我们把交互作用项纳入模型中来估计。我们的研究问题是是否[家庭作业]与反应变量[数学成绩]的关系在公立学校高于私立学校,如果是如此,我们即获得了另一个平等主义效果的研究结果。交互作用项的显著负向效果,将可以被解释成私立学校的效果比公立学校的效果弱。

为了在MLn软件中进行[家庭作业]与[公立]交互作用项的检验,我们必须以MULT指令来计算两个变量的乘积。下面的指令集显示交互作用项位于数据库的第19栏,在相乘指令之后,即跟随一个NAME指令来为这个新增变量命名(homepub)。要在模型中加入这一项,必须利用EXPL指令。在我们的模型中,交互作用项一律以固定系数来处理,表示我们不必以SETV指令来设定新的方差成分。以下的方框列出了MLn的语法:

```
MULT 'homew' 'public' C19
NAME C19 'homepub'
EXPL 'homepub'
```

分析结果如下:

$$[数学成绩] \approx \underset{(2.88)}{48.55} + \underset{(1.51)}{2.31} \times [家庭作业] - \underset{(3.55)}{3.29} \times [公立]$$

$$-0.50 \times [家庭 \times 公立]$$
$$(1.88)$$

变量与交互作用间的高相关的影响,可以在数值放大的[公立]系数的标准误上看到。同时[公立]变量的效果量并不大,导致[公立]变项无法达到统计的显著性。同样的现象也发生在[家庭作业]变量,但是它的效果量则增加(由1.98增加到2.31),而标准误将近是原来的两倍,也造成变量效果未达显著水平。由于交互作用项[作业×公立]也没有达到显著水平,因此我们可以为本模型的结果下一个结论:模型中没有任何一个变量具有显著的解释力。我们可以进一步利用本模型的模式拟合度与虚无模型的拟合度相比较,来检验上述说法是否正确。表4.10列出了偏差度与方差估计结果,本模型(模型7)比虚无模型的模式拟合度改善了166.01,虚无模型为3 800.78(表4.2),模型7为3 634.77(表4.10),此一可观的差值显示模型拟合获得明显改善,但这个结论与先前我们以个别固定效果系数的检验结果所得到的结论恰好相反。如果我们没有刚才的那些检验,或是我们没有拿这个模型的拟合度与虚无模型相比,或是我们没有检查方差成分的数值,我们不会有得出正面结论的机会,也就是说,我们没有机会宣称加入交互作用项对于反应变量变异量的解释具有意义,这个结论纯粹是由偏差度或方差成分数据所得出。到目前为止,我们尚未确切了解增加跨层级交互作用项是否造成模型的不稳定性,例如造成[公立]变量的主效果或交互作用效果的消失。

表4.10 模型7结果[译者分析8]

第二层		
参数	估计值	标准误
截距方差	56.21	19.09
斜率方差	16.30	5.40
斜率与截距协方差	-25.92	9.42
第一层		
参数	估计值	标准误
方差 Variance	53.34	3.47
偏差度 Deviance	3 634.77	

要找出显著的固定效果,最好的方法是把模型中的有关的变量移除,例如移除跨层级交互作用项(使我们回到模型6),或是移除某一个主要效果项。最好的选择仍是基于理论考虑或我们对于数据的理解,由于我们没有适当的理论依据,因此我们利用模型间拟合度的比较来决定拿掉哪一个部分。我们发现,如果模型没有[公立]与[作业×公立]两项,所得到的偏差度为3 639.04(表4.4的模型2结

果)，与目前的模型 7 偏差度的差值仅有 4.27，亦即两个模型没有差异，因而我们放弃有关[公立]与[作业×公立]两项的研究假设。

当我们进行跨层级交互作用的检验之时，多元共线性是造成解释问题的主因之一。由结果显示，这些彼此有关的变量的系数在不同模型间有明显的变动，但是他们的标准误变动的幅度更大。[公立]的主要效果从达到显著的 4.08(标准误为 1.90)，到具有一个与[公立]有关的跨层级交互作用模型后，则变成未显著的3.28(标准误为 3.55)。本节的分析结果并未支持我们的研究假设，即公私立学校对于学生的成绩表现具有显著的影响。不仅在主效果部分(对于截距的影响)没有发现，在跨层级交互作用(对于斜率的影响)也是如此。没有发现效果的原因，可能是真相就是这样，也可能因为我们样本数太少而效果亦微弱所造成，尤其是当组数偏低时(学校数少)或是总样本数偏低时，研究结果会因为多元共线性的存在而产生不稳定的结果(模型与模型间的差异很大)，模型估计的稳定性需要较大规模的样本(组数)来确保。

下面，我们将以 NELS88 的全体样本(21 580 学生而不是 519，1 003 所学校而不是 23)来进行分析，我们将使用相同的模型，亦即[家庭作业]、[公立]、[作业×公立]来预测[数学成绩]，分析的结果会发现所有的变量都有显著的效果，这与我们之前的不显著的结论是相冲突的。

4.3.5 NELS88 完整数据的分析

以 NELS88 的完整数据来进行分析所使用的模型是先前最后一个模型(模型 7)，在模型 7 当中，[公立]与[家庭作业]两个主效果与跨层级交互作用[作业×公立]之间具有高度相关。

如果我们以较大规模的数据库重新检验模型 7，效果会容易被显现出来。前面所使用的数据库是全体数据的一个样本，其中的多元共线性问题比整个全体数据库的多元共线性严重。在表 4.9 当中，列出了全体样本的变量相关，此时，较高的相关出现在跨层级交互作用项与两个主效果，分别是[公立]的 0.52 与[家庭作业]的 0.70。模型 7 重新分析的结果如下：

$$
\begin{aligned}
[\text{数学成绩}] \approx\ & \underset{(0.38)}{57.72} + \underset{(0.10)}{0.92} \times [\text{家庭作业}] - \underset{(0.43)}{6.05} \times [\text{公立}] \\
& - \underset{(0.12)}{0.68} \times [\text{家庭} \times \text{公立}]
\end{aligned}
$$

由结果可以发现，相同的模型以不同的数据库所分析得到的结果不同，前面的模型 7 的分析结果发现[公立]、[家庭作业]与跨层级交互作用[作业×公立]都没有显著的效果。但是，本节改以大型数据库来分析，每一个变量都有很强的效果与

很小的标准误,均具有统计的显著性。[家庭作业]是[数学成绩]显著的解释变量,[公立]变量对于[数学成绩]则有显著的负向效果,表示公立学校的成绩较差;跨层级交互作用也有显著的负向效果,表示私立学校有较强的[家庭作业]与[数学成绩]关联。

前一节的结论是我们的假设并没有获得支持,然而,本节的结果则相反,主效果(公私立的不同学校型态)与交互作用均具有显著的负向效果,两者均支持我们的研究假设,也就是私立学校比公立学校有较强的平等主义色彩(egalitarian)。效果的负向符号表示对于私立学校有利的结果(因为公立学校编码为1,私立学校编码为0)。

本模型的随机效果结果列于表4.11。在这个部分,我们要特别提醒以因果关系的口吻来解释公私立学校要特别小心。一般而言,观察性调查数据(不是从实验过程得到的数据)并没有因果结论的论证基础。能够用来解释公私立学校差异的变量有很多,在一个观察性调查研究中,所获得的结论是基于我们所选取的解释变量,或是在研究者控制某些变量之下所获得的结果。我们可以很轻易地证明,通过数据的操作(例如变量的中心化),会得到完全不同的结论。我们将在第5.2节讨论解释变量的组平减的影响。在第5.2节,我们以与本节相同的模型,但是变量则进行中心化处理,得到的结果又不一样了,公立学校的效果从负向变成正向,再次说明在随机或固定的回归模型中,如果存在着多元共线性,结果的解释要非常小心。

表4.11 以NELS 88完整数据得到的模型8结果[译者分析9]

第二层		
参数	估计值	标准误
截距方差	18.26	1.28
[家庭作业]斜率方差	0.44	0.10
斜率与截距协方差	-0.18	0.28
第一层		
参数	估计值	标准误
方差 Variance	71.72	0.72

4.3.6 消除[家庭作业]增加[白人]的小样本分析

前面我们以小数据库进行分析的结果,并没有确认两个学校层次变量[学校规模]与[公立]的效果,这两个变量无法解释[家庭作业]的截距与斜率的校间差异。在本节当中,我们将把重心放回学生特征,然后再把学校特征放进模型。下一个模型中,我们将检验学生的种族特征:白人(编码为1)或非白人(编码为0)(以[白

人]表示),是否会影响[数学成绩](有关变量的原始编码请参见附录)。在我们的数据中,白人学生为多数(占75%)。

从前面的模型继续往下延伸,我们消除了交互作用项[作业×公立],然后增加学生层次解释变量[白人],指令如下所示:

```
EXPL'homepub''white'
```

学校层次变量[公立]与学生层次变量[家庭作业]仍保留在模型中,此一模型(模型9)的分析结果如下:

$$[\text{数学成绩}] \approx \underset{(2.12)}{46.61} + \underset{(0.88)}{1.91} \times [\text{家庭作业}] + \underset{(0.96)}{3.36} \times [\text{白人}] - \underset{(1.72)}{3.91} \times [\text{公立}]$$

固定效果部分的结果显示,在[家庭作业]与[公立]被考虑的情况下,[白人]变量可以让[数学成绩]增加3.36分。[白人]与[公立]系数的方向相反,但强度相当。从这两个系数,可以得知在家庭作业时数相等的情况下,公立学校非白人学生的数学成绩的预测分数最低;私立学校中的白人学生数学成绩的预测分数最高。如果我们利用图形来说明,如图4.2,可以看到上述的结果。

图4.2中,公立与私立学校各有两条回归线,一条为白人学生,一条为非白人学生,私立学校的回归线趋势最强,可以从两方面来说明:第一,整体来看,私立学校的表现均高于公立学校学生;第二,私立学校白人学生比公立学校的两种学生的表现来得好,这是一种交互作用的结果。但是图表所揭示的结果是否可靠?基本上,两种不同类型学校的比较有些问题,因为[白人]变项系数的标准误较小,[公立]变项系数的标准误较大,表示[白人]变项系数的估计较为精确。因此我们必须注意如此的结论是否会得到错误的统计决策。

对于模型9的结果,可以从两个角度来解释我们的发现,第一是从个别的系数来看,第二是从整体模型拟合的角度来看。首先,[白人]与[公立]两个变量系数的强度除以标准误可得到两者的z检定值,[公立]的z值为2.27,[白人]的z值为3.5,虽然都达到统计显著水平,但是效果强度都不大,两个变量的系数效果都接近0,因为两者的标准误都偏高,显示他们是较不稳定的估计值。以图形来解释效果的系数强度,可能得到这是一个很稳定明确的效果的错误印象。另一方面,从模式拟合的角度来检验[白人]与[公立]两变量的效果,我们可以比较纳入此二变量的模型与没有纳入的模型(表4.4的模型2)的偏差度差值,本模型的偏差度与方差结果列于表4.12,偏差度差为15.79,显示模型拟合改善幅度并不十分明显。两个学校层次变量的效果虽然显著,但是效果小到我们不得不注意我们是否有过度结论,犯下第一类型错误。

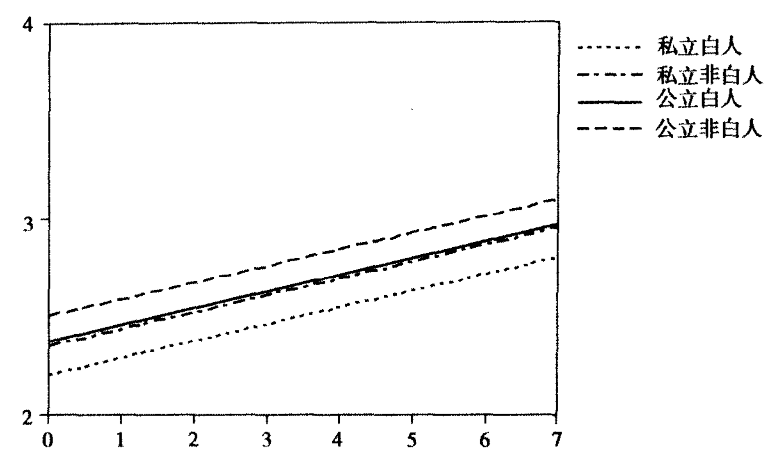


图 4.2 种族与学校类型对家庭作业预测效果的影响

表 4.12 模型 9 结果

第二层		
参数	估计值	标准误
截距方差	52.28	17.88
[家庭作业]斜率方差	15.83	5.25
斜率与截距协方差	-25.34	9.08
第一层		
参数	估计值	标准误
方差 Variance	52.64	3.42
偏差度 Deviance	3 623.25	

4.3.7 增加[白人]的随机部分

在下一个模型中,我们将把学生层次解释变量[白人]的系数增加随机的设定,此一变动只需要一个简单的指令,也就是把[白人]的方差设定在第二层:

```
SETV 2 'white'
```

此一变动的结果,对于固定部分与随机部分的影响会远超过我们的预期。固定效果的结果如下:

$$[\text{数学成绩}] \approx \underset{(2.26)}{48.18} + \underset{(0.88)}{1.95} \times [\text{家庭作业}] + \underset{(1.50)}{2.67} \times [\text{白人}] - \underset{(1.58)}{4.94} \times [\text{公立}]$$

结果显示[家庭作业]与[公立]的固定效果与前面模型的估计结果相近,但是[白人]变量的估计值就有明显的差别,z 检定值只有 1.78,未达统计显著水平。将参数设定成可自由变动的随机系数,会增加模型的不稳定性,我们可以从[白人]变量系数标准误的提高看到此一现象,前一个模型的标准误为 0.96,到了本模型(模型 10),提高到 1.5。我们在稍早的章节中也曾经看到,如果把[家庭作业]由固定改为随机,固定效果的显著水平产生明显变化(见表 4.3 与表 4.4 的比较)。

模型中加入随机斜率的代价,就是固定效果的参数估计值会变得不稳定。在这种情况下,如何来处理[白人]? 是否设定为随机或固定? 这个决策就变得非常关键。最好的决定因素还是理论。例如,我们可以理解数学成绩并不是完全是个别层次学生的效果(例如白人或非白人),主要还是学校特征的影响结果。如果基于这样的理论观点,那么本模型(模型 10)的设定就较为合理。不管怎么说,前面的两个模型非常相似,但是却出现了固定效果的结论不一致的现象,显示理论的导引是一个重要的因素。在本模型中,随机系数的设定所造成的影响,在表 4.13 所呈现的随机部分的结果中也可以明显地看到。

当我们在模型中增加一个随机斜率,会创造三个新的方差成分:一个[白人]斜率方差,两个协方差。两个协方差是[白人]与[家庭作业]的协方差及[白人]的斜率与截距的协方差。由表 4.13 可知,[白人]变量的系数方差并没有达到统计显著水平,方差为 24.22,标准误为 14.13,z 值为 1.71。我们的研究假设并没有获得支持,也就是学校特征不会造成白人与非白人学生在数学成绩上的差异。

但是,我们还是必须提醒上述的结论必须小心为之,因为多元共线性问题依旧还在困扰着我们。表 4.13 的方差成分与表 4.12 相比,我们会发现一些特别的现象,其中之一是第二层变异成分的加入导致组内方差减小,从表 4.12 的 52.64 降至本模型的 51.15,我们并没有预期到此变化,因为各阶层的误差项是假设为无相关(有关误差变异的特性与彼此间关联的细节讨论请参见第 5.3 节)。此外,截距方差的变化也很可观,从 52.28 增加到 64.41,标准误也增加。此变化是由于变量间的相关所导致,我们可以从协方差来观察变项间所存在的相关。由此可知,我们已经无法就这些估计值的表面数值来解释这些参数的意义。

另外,从技术的角度来决定随机斜率是否值得放入模型,是检查模型的拟合度。本模型的偏差度为 3 618.85(见表 4.13),拟合度改善的幅度是 4.40,损失的自由度为 3,并未达到统计显著水平,显示增加随机斜率并无法有效改善模型的拟合性。

表4.13 模型10结果

第二层		
参数	估计值	标准误
截距方差	64.41	25.86
[家庭作业]斜率方差	15.68	5.20
[白人]斜率方差	24.22	14.13
[家庭作业]与截距协方差	-26.97	10.41
[白人]与截距协方差	-20.27	15.40
斜率与截距协方差	2.75	6.31
第一层		
参数	估计值	标准误
方差 Variance	51.15	3.37
偏差度 Deviance	3 618.85	

在本章的第一部分,我们发现带有随机斜率的[家庭作业]模型的拟合度改善十分显著,但是对于[白人]则不然。如果我们可以找到适当的理论基础来告诉我们要如何决定[家庭作业]与[白人]变量的随机斜率的放或不放,事情或许比较单纯,然而正因为现在我们没有理论依据,因此我们就从数据的角度来决定要把[白人]系数的方差设定为0(也就是[白人]变量的斜率为非随机的固定值)。

值得注意的是,我们并不认为把[白人]的斜率设为固定是一个正确的决定,如果方程式中增加其他变量或移除某些变量,或许[白人]变量的斜率的随机效果会变成显著也不一定。

4.3.8 [白人]斜率设为固定增加[平均SES]

在下一节的多层次分析的示范中,我们将创造一个学校层次的聚合变量[平均SES](meanses)。产生方式是用各校的学生层次变量[社经地位]求出各校平均数,即为聚合变量(aggregated variable)。在MLn中,可以利用MLAV指令来进行变量的聚合。然后将这个变量放入MLn数据库中,成为C20变量,并命名为meanses。

```
MLAV 'school' 'ses' C20
NAME C20 'meanses'
```

我们首先检验[平均SES]变量的效果。如果[平均SES]具有显著的影响力,将会在截距项发现一个显著降低校间方差,然后,可以进一步地检验[平均SES]是否可以解释模型当中其他的显著校间变异(也就是[家庭作业]系数的变异)。下

面的 MLn 语法清除了模型中的[白人]系数的方差,并新增[平均 SES]。

```
CLRV 2 'white'
EXPL 'meanses'
```

在进行分析之前,我们必须先指出我们的数据库中具有多元共线性的问题。[公立]与学校的平均社经地位两个学校层次变量具有高相关,小样本数据库所求出的相关为 -0.78,完整的 NELS88 所计算出来的相关为 -0.54(见表 4.1)。相关系数的负值意味着公立学校比私立学校平均来说,学生的社经地位较低。这个相关系数会影响参数的估计。以我们的范例数据为例,[公立]变量的系数跌至接近 0 的数值,而且不再具有统计显著性,[平均 SES]变量系数的 z 检定值为 2.76,达统计显著水平。新的模型估计结果如下:

$$
\begin{aligned}
[\text{数学成绩}] \approx\ & \underset{(2.14)}{44.58} + \underset{(0.87)}{1.93} \times [\text{家庭作业}] + \underset{(0.95)}{3.14} \times [\text{白人}] + \underset{(2.12)}{0.17} \times [\text{公立}] \\
& + \underset{(1.82)}{5.03} \times [\text{平均 SES}]
\end{aligned}
$$

现在,重要的问题在于校间的差异是由[公立]还是[平均 SES]所解释?再一次的,这个问题的回答还是要先看理论怎么说,例如有关社经地位对数学成绩的影响的相关文献。当然我们也可以利用偏差度来判定两个变量的影响力,基于模型拟合度,我们可以决定究竟是[公立]变量解释得比较多,还是[平均 SES]解释得比较多。

本模型(模型 11)的偏差度与方差成分估计结果列于表 4.14,模型 11 比起模型 5(没有[平均 SES]的模型,结果列于表 4.12),模型拟合显著获得改善,偏差度差为 3 623.25 - 3 616.83 = 6.42,自由度为 1。

表 4.14 模型 11 结果

第二层		
参数	估计值	标准误
截距方差	50.11	17.20
[家庭作业]斜率方差	15.43	5.10
斜率与截距协方差	-25.50	8.92
第一层		
参数	估计值	标准误
方差 Variance	52.72	3.43
偏差度 Deviance	3 616.83	

4.3.9 移除学校特征[公立]变量

由于我们已经知道[SES]对于学生学习成果有影响。因此我们保留[平均SES],移除[公立]变量,把仅带有[平均SES]的模型的偏差度与仅带有[公立]的模型的偏差度相比,可以得知两个模型的优劣。我们以EXPL指令来进行设定。

```
EXPL 'public'
```

固定效果的分析结果如下：

$$[\text{数学成绩}] \approx \underset{(1.74)}{44.68} + \underset{(0.87)}{1.93} \times [\text{家庭作业}] + \underset{(0.95)}{3.15} \times [\text{白人}] + \underset{(1.28)}{4.93} \times [\text{平均SES}]$$

从上述结果发现,移除了[公立]变量后,[平均SES]系数的标准误比前一个模型降低。此时,当学生都有相同的做作业时数时,社经地位最高的白人学生,数学成绩预测分数最高。

表4.15列出了模型12的随机效果与模型偏差度数据。偏差度显示[公立]变量的移除并没有让模型变得比较糟糕。表4.15的偏差度为3 616.83,与前一个模型相当(表4.14的模型11)。也就是说,[公立]变量的移除不改变模型拟合度。若把模型9(表4.12)与模型12(表4.15)的结果相比,偏差度分别为3 623.25与3 616.83,显示带有[平均SES]的模型拟合度则比带有[公立]变量的模型理想。由于[平均SES]的模型具有较理想的拟合度,因此我们继续利用这个变量来进行分析。

表4.15 模型12结果

第二层		
参数	估计值	标准误
截距方差	50.01	17.20
[家庭作业]斜率方差	15.42	5.10
斜率与截距协方差	-25.49	8.92
第一层		
参数	估计值	标准误
方差 Variance	52.72	3.43
偏差度 Deviance	3 616.83	

再一次,我们要提醒模型优劣的选择需尽可能地基于理论文献,而非技术指

标。以我们的例子来说,全校平均社经地位的高低确实是比学校类型更理想的解释数学成绩的学校特征。以下,我们将建立一个模型,带有[平均 SES]以及与[家庭作业]的跨层级交互作用项。跨层级交互作用项的纳入意味着[平均 SES]将被用来解释[家庭作业]斜率的显著组间变异。

4.3.10　增加[家庭作业]与[平均 SES]的交互作用

前面的分析中,我们获得一个结论,[家庭作业]的斜率具有显著的随机变化特性。但是我们并无法解释是什么原因造成此一学校间的变异。[学校规模]或[公立]两变量与[家庭作业]的交互作用均未达到统计显著水平。在本节当中,我们以聚合得来的[平均 SES]变量来"解释"[家庭作业]斜率的组间变异。首先,我们必须创造一个跨层级交互作用项,可利用 MLn 的计算指令 CALC 来进行这项工作(在第 4.3.4 节是利用相乘指令 MULT,两个指令都可以用来计算交互作用项)。新的跨层级交互作用项命名为[家庭 × MSES],再利用 EXPL 指定此一变量名称为 homeses。

```
CALC C21 = 'homew' * 'meanses'
NAME C21 'homeses'
EXPL 'homeses'
```

固定效果的分析结果如下:

$$[\text{数学成绩}] \approx \underset{(1.76)}{44.58} + \underset{(0.88)}{1.99} \times [\text{家庭作业}] + \underset{(0.95)}{3.15} \times [\text{白人}] + \underset{(2.88)}{3.99} \times [\text{平均 SES}] + \underset{(1.56)}{0.57} \times [\text{家庭} \times \text{MSES}]$$

上述的分析结果并不令人满意,因为跨层级交互作用并不显著,而且模型拟合度也没有改善,甚至于从数据表面来看,模型反而变差,[平均 SES]的标准误从 1.28提高到 2.88,系数的数值反而降低。[平均 SES]的主要效果从前面的 $z=3.88$ 的显著结果变成不显著的 $z=1.40$。根据此一结果,我们的结论是:[平均 SES]不仅没有主要效果,也没有跨层级交互作用。但是此一结论与我们先前的结论产生冲突。模型 8 发现,[平均 SES]的主要效果与交互作用[家庭 × MSES]的高相关($r=0.83$,参见表 4.9)导致模型十分不稳定。

表 4.16 报告了模型 13 的偏差度与方差成分。与表 4.15 的偏差度相比,两个模型的偏差度差值为 0.13,增加跨层级交互作用并没有显著改善模式拟合度。反倒是因为增加了跨层级交互作用,我们付出了[平均 SES]的个别参数估计值趋向

不稳定的代价(标准误变大)。增加一个与其他变量有相关的变量(本节的例子是跨层级交互作用项),增加参数估计的不稳定性,而且在小样本的多层次分析,多元共线性的影响更为明显。

表 4.16 模型 13 结果

第二层		
参数	估计值	标准误
截距方差	49.90	17.13
[家庭作业]斜率方差	15.28	5.06
斜率与截距协方差	-25.32	8.86
第一层		
参数	估计值	标准误
方差 Variance	52.73	3.43
偏差度 Deviance	3 616.70	

4.3.11 增加另一个学生层次变量

在学校效能的研究文献中,社经地位扮演一个重要的角色,高社经地位的学生平均而言会有较佳的学业成绩表现,已是公认的结论。对学校来说,此一结论也成立,亦即社经地位较高的学校,平均来说,学生的成绩较佳。在前面的分析中,我们把[平均 SES]放入模型中,在本节当中,我们则把学生层次的[SES]变量也放入模型。

我们的新模型有三个学生层次变量,[家庭作业]、[白人]、[SES],以及一个学校层次变量[平均 SES],但没有跨层级交互作用。MLn 的语法当中,我们以 EXPL 指令移除了交互作用项[家庭 × MSES],增加[SES]。

```
EXPL 'homeses' 'ses'
```

固定效果的分析结果如下:

$$[\text{数学成绩}] \approx \underset{(1.71)}{45.65} + \underset{(0.83)}{1.83} \times [\text{家庭作业}] + \underset{(0.96)}{2.22} \times [\text{白人}] + \underset{(0.53)}{2.21} \times [\text{SES}] + \underset{(1.37)}{2.97} \times [\text{平均 SES}]$$

在本模型中(模型 14),所有的系数的 z 检定值接近 2.00,除了学生层次的[SES]变量,其 z 值达 4.17。增加学生层次的解释变量[SES]如预期般降低了学校层次解释变量[平均 SES]的估计值(表 4.1 中的 $r=0.70$)。比较参数的数值,我们

可以发现[平均 SES]的强度从模型 12 的 4.93 降低至模型 14 的 2.97,标准误则由模型 12 的 1.28 变成 1.37。

表 4.17 列出了模型 14 的偏差度,与表 4.15 的偏差度相比,损失了一个自由度,但偏差度降了 16.75,显示模型的拟合获得显著改善。表 4.17 的组间与组内方差都有些微降低,是增加学校层级社经地位[平均 SES]的影响。

表 4.17 模型 14 结果

第二层		
参数	估计值	标准误
截距方差	46.60	16.10
[家庭作业]斜率方差	13.78	4.60
斜率与截距协方差	-23.02	8.16
第一层		
参数	估计值	标准误
方差 Variance	51.12	3.32
偏差度 Deviance	3 600.08	

如果使用较大的数据库,模型受到多元共线性的影响就比较小。另一个优点是在大样本下跨层级交互作用效果的检定力较大。以下,我们就以 NELS88 的完整样本来进行分析。

4.3.12 NELS88 完整数据库分析

本节以 NELS88 的完整数据来分析前一个模型(模型 14)。此一模型有三个具有高相关的解释变量:[家庭作业]、[SES]、[平均 SES](相关系数列于表 4.1),尤其是[SES]与[平均 SES]的相关达 0.65。以大型数据所分析的结果如下:

$$
\begin{aligned}
[\text{数学成绩}] \approx\ & \underset{(0.71)}{47.61} + \underset{(0.05)}{1.22} \times [\text{家庭作业}] + \underset{(0.15)}{1.78} \times [\text{白人}] + \underset{(0.09)}{3.73} \times [\text{SES}] + \\
& \underset{(0.21)}{4.00} \times [\text{平均 SES}]
\end{aligned}
$$

本模型以大型数据所得到的固定效果结果与前一个模型差异甚大,有几个理由:第一,前面所使用的小样本是 NELS88 内的一个子样本,而且不是随机抽取,因此小样本数据所得到的结果未必有代表性,系数数值大小也可能会有明显的不同;第二,由于 NELS88 的全体样本规模庞大,在两个层次都有庞大数量的观察值,标准误很小,导致每一个参数的显著性检验结果都产生变化。例如[家庭作业]斜率

的 z 值为 24.4,但是在小样本时只有 2.20,同样,[白人]变量的斜率的 z 值为 11.87,小样本时仅有 2.31,[SES]的 z 值为 19.05,小样本则为 2.17,[平均 SES]的 z 值为 19.08,在小样本仅为 2.17。因此,改用大样本的数据来进行分析,会使得原本低迷的显著性检验结果变得十分可观,显示大样本会提高统计的检定力。

表 4.18 NELS 88 **所有数据的分析结果(模型** 15)

第二层		
参数	估计值	标准误
截距方差	7.39	0.74
[家庭作业]斜率方差	0.35	0.08
斜率与截距协方差	-0.79	0.22
第一层		
参数	估计值	标准误
方差 Variance	66.12	0.66

4.4 第三部分

4.4.1 以社经地位为解释变量

在第三部分,我们将检验学生所处的环境对于数学成绩的影响。所谓"环境"是从两个层面来定义:个体的家庭环境和学校的环境。个体家庭环境以父母的社经地位[SES]来衡量,该变量是合并自教育水平与薪资所得。而作为学校环境的变量,我们采用了学校中少数民族学生所占的比例[种族比例],以及平均社经地位[平均 SES]。

我们开启一个新的 MLn 工作表,定义反应变量(RESP 指令)为[数学成绩],解释变量(EXPL 指令)为截距(cons)与[SES],加上一个第一层方差与第二层的截距方差。整组指令如下:

```
RESP 'math'
EXPL 'cons' 'ses'
SETV 1 'cons'
SETV 2 'cons'
```

从语法内容可知，这个模型仅有一个学生层次解释变量[SES]，学生与学校层次的方差只有一个：截距的随机系数。模型分析结果如下：

$$[\text{数学成绩}] \approx \underset{(0.83)}{51.20} + \underset{(0.56)}{4.35} \times [\text{SES}]$$

结果显示，[SES]的固定系数具有非常显著的效果，z 检定值为7.77，如果整体模型的拟合有所改善，那么[SES]这个参数的效果就可以被认真地接受。同样，模型拟合改善幅度的评估是由带有[SES]变量的模型与第一部分所报告过的虚无模型的偏差度相减而得。

本模型（模型16）的偏差度数据列于表4.19，两个模型的偏差度差值为52.42（模型16的偏差度为3 748.36，虚无模型的偏差度为3 800.78），此一差值非常大，因此我们假设：学生的家庭环境对于数学成绩具有影响力。

表4.19 模型16结果[译者分析10]

参数	估计值	标准误
阶层二方差	11.80	4.61
阶层一方差	75.20	4.77
组内相关 ICC	0.22	
偏差度 Deviance	3 718.29	

与虚无模型相比，模型16的方差成分数值也有所改变。截距的组间方差从虚无模型的24.85降至11.80，减少了53%。以我们对可解释变异的定义来说，R_B^2 为0.53。组内方差也降低了，从虚无模型的81.24降至75.20，减少了7%，因此 R_W^2 为0.07。基于此一分析结果，我们发现[SES]的效果主要是学校间的效果，而非组内的效果（个体间的差异）。从这个现象来看，我们可以预期[SES]的系数在各校之间显著不同。有关多层次模型当中的组间与组内可解释变异量的计算与意义，我们将在第5.3节中进行详细的讨论。

4.4.2 增加随机斜率

本节的示范是在模型中增加一个随机斜率，背后的假设是[SES]与[数学成绩]的关系在各校可能不一样，例如老师的不同教学风格导致关系的不同。在一个平等主义取向的老师身上，老师会设定一个让所有的学生都可以达到的目标，不论学生的背景为何，最后大家的学习成果会接近相同，精熟度相同（精熟学习）。其他类型的教学风格就可能大不相同，学生的既存个别差异可能会深深影响学习的成

果,例如按能力分班的影响。因此,教学方法的差异可能导致[SES]与[数学成绩]存在不同的关系模式。

在模型中增加[SES]的随机斜率,只需要一个 MLn 指令,在第二层设定[SES]斜率方差。

```
SETV 2 'ses'
```

固定效果的结果与先前的模型很类似:

$$[\text{数学成绩}] \approx \underset{(0.83)}{51.20} + \underset{(0.56)}{4.35} \times [\text{SES}]$$

从前面的分析中,我们知道增加随机斜率会改变固定效果的强度,也可能改变系数的标准误,第一部分的解释变量[家庭作业]与第二部分的解释变量[白人]都发生这个现象。这两个例子都是原来的高度显著的系数,在增加随机斜率之后,却都变成接近 0 的不显著的结果。但是在本范例中,[SES]的效果则无改变。

表 4.20 模型 17 结果[译者分析11]

第二层		
参数	估计值	标准误
截距方差	11.80	4.61
[家庭作业]斜率方差	0	*
斜率与截距协方差	0	*
第一层		
参数	估计值	标准误
方差 Variance	75.20	4.77
偏差度 Deviance	3 748.36	

从表 4.20 当中,可以看到为何[SES]的固定效果不受随机斜率纳入模型的影响。记住:增加一个随机斜率,模型中会增加两个参数,一个是斜率方差,一个是斜率与随机截距的协方差。表 4.20 当中最特别的是第二层的方差成分中,[SES]的斜率方差为 0,协方差也为 0。表中所有的数据都与前一个没有随机斜率的模型相同,截距方差相同,第一层的方差和偏差度都相同。换言之,此模型(模型 17)与前一个模型(模型 16)是完全相同的模型,因为并没有任何的[SES]随机斜率被估计到。如果方差为 0,在 MLn 软件,方差的估计并不会收敛到 0,当方差接近 0 时,迭代估计将会缓慢下来,因此,在 MLn 当中,对于方差为 0 的情况,会把方差设定为

接近 0 的数值，以提高迭代的速度，然后产生一个不显著的方差估计值。其他的软件对于这个问题有不同的处理方法。

前面的分析结果并不支持“某些学校的[SES]与[数学成绩]关系有所不同”这个假设。

4.4.3　增加[种族比例]

下面我们将增加学校特征[种族比例]（minority）来进行分析，但是把[SES]的斜率改回固定。指令如下，

```
CLRV 2 'ses'
EXPL 'minority'
```

分析的结果如下：

$$[\text{数学成绩}] \approx \underset{(1.13)}{53.12} + \underset{(0.56)}{4.34} \times [\text{SES}] - \underset{(0.35)}{0.80} \times [\text{种族比例}]$$

分析结果显示，[SES]的系数估计值与前一个模型相似，不论是系数的强度或标准误，z 检定值为 7.75，达到显著水平。偏差度与随机部分的结果列于表4.21，[种族比例]的影响列于第二层的变异成分中。一般来说，增加变量可能造成两个层次方差的估计的变化，例如增加一个第一层变量可能使得两层的方差都降低，我们可以预期学生层次变量影响的组内变异成分较大，因此新增学生层次变量会对第一层的方差影响较大。

表 4.21　模型 18 结果[译者分析12]

第二层		
参数	估计值	标准误
截距方差	9.43	3.90
第一层		
参数	估计值	标准误
方差 Variance	75.02	4.76
偏差度 Deviance	3 743.39	

到目前为止，[SES]与[家庭作业]这两个学生层次变量在学校间都有很大的变异。因此这两个变项加入模型后，会造成组间变异的明显减低。但是，增加学校

层次变量仅会影响组间变异,因为学校层次变量的定义是仅有学校间的效果而没有组内效果。我们预期在加入[种族比例]变量之后,本模型的组间方差(截距项的)会比没有增加的模型变小。

分析结果正如我们的预期,截距方差从表 4.19 的 11.8 降到 9.43,减少2.37,亦即组间部分增加 9.5% 的被解释变异量。如果要让学生层次变量仅影响组内变异,可以使用组平减程序,改变变量的数据形式,有关中心化的影响与解释将在第 5.2 节中讨论。

本模型(模型 18)的偏差度(见表 4.21)与前一个没有[种族比例]的模型(模型 17,见表 4.20)相比,减少了 4.97,显示增加了[种族比例]变量导致了模型拟合的稍许改善,以及 9.5% 的 R_B^2 的增加。

在模型比较过程中,我们并没有比较第一层的组内变异,因为我们假设第一层的组内变异不受新增的学校层次变量的影响。在多层次分析中,第一层与第二层残差是独立无关的。第二层的变量只有组间变异部分,不像第一层变量有组间与组内变异两个部分。在下一节的分析当中,我们将增加另一个学校环境变量[平均 SES],从数据当中,我们将证明模型拟合的改善(R^2 的增加)与个别参数的估计会有相冲突的结果。我们认为,模型整体拟合(total model fit)是衡量模型的改善程度以及个别参数估计的稳定度(信度)的最佳方法。

4.4.4 增加[平均 SES]

下一个模型是在前面的模型中,增加一个学校环境有关的变量[平均 SES]所延伸出来的新模型。

```
EXPL 'meanses'
```

在我们所使用的小型数据库中,[平均 SES]与学校层次变量[种族比例]几乎没有关系,相关系数为 -0.08(见表 4.1)。但是在 NELS88 的完整样本中,两者的相关较高($r=-0.36$),显示当学校少数民族学生较多时,学校的整体社经地位较低。当模型中同时具有两个学校环境变量时,估计的结果如下:

$$[\text{数学成绩}] \approx \underset{(1.03)}{53.08} + \underset{(0.61)}{3.89} \times [\text{SES}] - \underset{(0.32)}{0.68} \times [\text{种族比例}] + \underset{(1.39)}{2.86}[\text{平均 SES}]$$

在此模型中(模型 19),[平均 SES]的系数几乎是不显著的,z 检定值为 2.06,而[种族比例]则几乎没有变动,z 检定值为 2.13。表 4.22 列出了模型 19 的偏差度(3 739.61),与最初的模型(没有[种族比例]与[平均 SES]的模型 17)相比,偏差度差值为 8.75,改变的幅度并不大。若与前一个模型(仅有[种族比例],没有

[平均 SES]的模型18)的 3 743.39 相比,偏差度差值只有 3.78,改善幅度更小,显示新增的[平均 SES]变量对于模型拟合度并没有明显的影响。

从组间方差(截距方差)的变化来看也很有意思,当新增了两个学校层次变量后,组间方差降低了。若以被解释变异量的术语来说,两个变量"解释"了18%的组间变异。此一数值可能会被研究者解释为两个变量具有强大的效果。

表 4.22 模型19结果[译者分析13]

第二层		
参数	估计值	标准误
截距方差	7.23	3.23
第一层		
参数	估计值	标准误
方差 Variance	75.08	4.76
偏差度 Deviance	3 739.61	

表4.19的组间变异为11.80,表4.21的组间变异降低了9%,到了表4.22降低至7.23, R_B^2 又降了9%。如果从虚无模型开始,计算到表4.22的组间与组内变异的总解释方差 R^2 ,组间变异从虚无模型的24.85降至7.23,减少17.62,亦即 $R_B^2 = 0.71$,[SES]解释了53%的组间变异,两个学校特征变量解释了18%的组间变异。在组内变异部分,虚无模型为81.24,最后一个模型为75.08,差值为6.16, $R_W^2 = 0.08$。

在此我们所要强调的是,[平均 SES]的个别参数估计值与被解释变异量的增加,两者是矛盾的结果。对于此一矛盾,研究者应使用模型拟合的数据来判断最后的结论为何,整体模型拟合是用来评估个别参数估计量与可解释变异量的较为稳定可靠的指标。

为了了解不同数据库对于模型估计的差别,下面我们将以完整的NELS88数据库来分析前面的最后两个模型。我们将会看到:

- [SES]斜率方差不再被设为0,对于此系数的解释并没改变。
- [种族比例]系数是非常显著,而不是几无显著。
- 没有[种族比例]变量的模型比有[种族比例]变量的模型的偏差度小很多。

4.4.5 NELS88 完整数据库分析

首先我们分析的例子是以[SES]为解释变量,但系数设为随机的模型,也就是

模型 17。我们以大型数据库来重新分析[SES]与[数学成绩]关系的学校差异(模型 20),前面的模型(模型 17)无法得到与原来研究者(Raudenbush & Bryk, 1986)相同的结论,主要的原因之一应是样本过小且没有代表性。前面所使用的样本是 23 个学校,完整的 NELS 数据库则有 1 003 所学校,如此庞大的数据库将可以提高统计检定力(参见第 5.4 节的讨论)。

以完整数据库来进行分析,我们的预期是前面所得到的[SES]斜率方差与协方差为 0 的情况将会改变,会成为非 0 且显著的结果。模型 20 的分析结果如下:

$$[\text{数学成绩}] \approx \underset{(0.12)}{50.96} + \underset{(0.10)}{4.82} \times [\text{SES}]$$

表 4.23 模型 20 结果[译者分析14]

第二层		
参数	估计值	标准误
截距方差	11.11	0.66
斜率方差	0.54	0.35
斜率与截距协方差	1.68	0.33
第一层		
参数	估计值	标准误
方差 Variance	69.76	0.70
偏差度 Deviance	154 336	

从结果的数据可以发现,大型数据库所估计得到的各系数的标准误都变小,我们最关心的是是否[SES]的斜率在各校之间具有显著的变异,方差成分的结果列于表 4.23。斜率方差的数值除以标准误得到 z 检定值为 1.54,仍然不显著,但是这一次 MLn 报告了[SES]的方差(0.54)与协方差(1.68)估计值与标准误,但是[SES]与[数学成绩]关系具有校间差异的假设并没有被我们的观察数据所支持。

为了解释的目的,我们也把模型 18 重新以大型数据库来进行分析,此时模型中增加了[种族比例],同时把[SES]的斜率方差设为 0。语法如下:

```
CLRV 2 'ses'
EXPL 'minority'
```

分析的结果如下:

$$[数学成绩] \approx \underset{(0.19)}{53.16} + \underset{(0.09)}{4.76} \times [SES] - \underset{(0.05)}{0.71} \times [种族比例]$$

把本模型(模型 21)的结果与表 4.21 的模型 18 结果相比较,我们可以发现标准误都明显降低了许多。学校层次的[种族比例]在模型 18 当中呈现几乎不显著的结果(z 检定值为 2.29),但是到了本模型,则呈现高度显著(z 检定值为 14.2)。本模型只有一个随机截距,第一层截距与第二层方差的数据列于表 4.24。同样地,我们可以发现方差的标准误也降低了,使得显著性检验的结果更加提高。另一方面,偏差度的比较发现,本模型的偏差度为 154 184,与先前以同样数据库估计的模型相比之下缩小很多,自由度为 152,表示模型拟合度更为理想。

表 4.24 模型 21 结果[译者分析14]

第二层		
参数	估计值	标准误
截距方差	8.78	0.55
第一层		
参数	估计值	标准误
方差 Variance	69.93	0.69
偏差度 Deviance	154 184	

对于解释变异量 R^2 的估计,由于数据库不同,因此我们需要重新计算虚无模型的方差成分估计值,得到的结果组内方差为 76.62,组间方差为 26.56。表 4.24 的组间变异为 8.78,降低了 17.78,因此 $R_B^2=0.67$;组内变异则降为69.93,降低了 6.69,得到 $R_W^2=0.09$。虽然我们以两个大小不同的数据库来分别进行估计,而且小数据库也不是从大数据库中随机抽样取得,但是得到的解释变异量却很接近。得到的结论相似,亦即[SES]与[种族比例]两个变量合在一起可以解释组间变异的很大一部分,而[SES]变量则仅可以解释组内变异的一小部分。

对于学校环境的影响,我们的分析结果发现,以学生层次变量[SES]与学校层次变量[种族比例]与[平均 SES]这三个解释变量来进行多层次分析,发现了环境因素确实对于[数学成绩]有一些影响。但是为何会得到不显著的[SES]系数方差,则还是一个谜。我们将在第 5.2 节讨论[SES]系数的随机与非随机问题。在该节中,我们所关心的问题是[SES]变量的中心化对于斜率方差的影响,如果把[SES]变量减去学校平均数(组平减)得到中心化变数,[SES]的斜率的校间差异将会变得非常显著,这与我们以原始型态的[SES]变量所得到结果有很大的不同。

4.5 第四部分

4.5.1 以班级规模与跨层级交互作用所进行的分析

前面以小型数据库所进行的各模型分析的结果中,我们都还没有得到显著的跨层级交互作用的结果。现在,我们有了更大的数据库,以及更多的变量,因此我们另外挑选一个学校层次解释变量[生师比](学生数除以老师数的 ratio,反应班级规模)来进行分析。较高的[生师比]意味着班级规模较大。在加州,耗费数百万美元进行班级规模缩减,企图改善公立学校低成就学生的表现的教改经验发现,班级规模的确是影响学生学习成就的重要决定因素之一。如果较低的[生师比]可以预测出较高的[数学成绩],应可得到显著的负向回归系数。

班级规模的效果可以从两方面来衡量,首先是整体效果,其次是交互效果。交互效果是指[家庭作业]与[生师比]的交互效果,背后的假设是:当班级规模较小时,学生做作业时间比较少;当班级规模较大时,学生要做较多作业。如此的假设意味着我们将会得到一个正向的交互作用效果,[家庭作业]与[数学成绩]的关系会因为班级较大而变得更强。

我们重新启动 MLn 软件,重新定义变量与模型,反应变量(RESP 指令)为[数学成绩],解释变量(EXPL 指令)为截距(cons)、[家庭作业]与[生师比]。方差部分,我们定义第一层与第二层的截距方差,以及[家庭作业]在第二层具有方差(随机)。整组指令如下:

```
RESP 'math'
EXPL 'cons' 'homew' 'ratio'
SETV 1 'cons'
SETV 2 'cons' 'homew'
```

分析结果如下:

$$[\text{数学成绩}] \approx \underset{(3.92)}{47.97} + \underset{(0.91)}{1.99} \times [\text{家庭作业}] - \underset{(0.20)}{0.10} \times [\text{生师比}]$$

此一结果并没有支持我们的假设,因为[生师比]系数并不显著。然而,当变项之间具有相关时,以个别系数的估计结果来进行解释可能会有所误导。虽然两个解释变量[家庭作业]与[生师比]的相关并不高($r=0.16$),但是我们还是要小心。

我们应该从模型拟合的角度来检验[生师比]的主要效果是否具有意义。我们把本模型(模型22)的偏差度(3 639.04)与前面的模型2(表4.4)的偏差度(3 638.82)相比,得到偏差度差值非常小,不到1,因此我们可以确实得到[生师比]对于数学成绩没有解释力的结论(以小样本进行分析的结果)。

表4.25 模型22结果[译者分析15]

第二层		
参数	估计值	标准误
截距方差	59.30	20.01
斜率方差	16.80	5.55
斜率与截距协方差	-26.25	9.67
第一层		
参数	估计值	标准误
方差 Variance	53.30	3.47
偏差度 Deviance	3 638.82	

4.5.2 [生师比]与[家庭作业]的交互作用

前面分析中,我们的假设并没有获得支持,以下,我们进行另一个也与[生师比]有关的假设检验。我们的假设是数学成绩与家庭作业时数的关系在小班的情况下强度较弱,在这种情况下,我们预期需要做很多作业的学生在小班的情况下受益最大。用统计的术语来说,我们会说[生师比]具有交互作用。如果假设成立,我们应该可以得到一个显著的[家庭作业]与[生师比]跨层级交互作用。

利用MLn软件的MULT指令,可以把[家庭作业]与[生师比]两个变量相乘来创造出跨层级交互作用变量,然后把这个新变量放进数据库(第C22变量),用NAME指令命名为homratio([作业×生师比]),语法如下:

```
MULT 'homew' 'ratio' C22
NAME C22 'homratio'
EXPL 'ratio' 'homratio'
```

模型分析结果如下:

$$[\text{数学成绩}] \approx \underset{(1.72)}{46.32} + \underset{(2.06)}{2.91} \times [\text{家庭作业}] - \underset{(0.11)}{0.05} \times [\text{作业} \times \text{生师比}]$$

再一次地,由于跨层级交互作用项的系数未达统计显著水平,分析结果并没有支持我们的假设。若把本模型(模型23)的偏差度3 638.79与仅带有[家庭作业]的模型的偏差度3 639.04(模型2,表4.4)相比,偏差度差值也非常小,模型拟合度没有任何改善。我们观察最后两个模型的固定部分,可以发现跨层级交互作用项的增加,由于变项间存在着相关,导致参数估计出现不稳定的现象。模型24的[家庭作业]与[作业×生师比]相关达0.86(表4.9),由于这个高相关的存在,使得在前一个模型(模型23)几乎不显著的[家庭作业]系数(估计值为1.99,标准误为0.91),到了本模型就真的不显著了,虽然估计值增高到2.91,但是标准误更提高到2.06,z检定值为1.41。

再一次地,如果改用NELS88完整数据库来对相同的模型进行分析,我们得到了不同的结论。当我们的样本扩大之后,跨层级交互作用就比较容易被检测出来(有关统计检定力的讨论请参见第5.4节)。

表4.26 模型23结果[译者分析15]

第二层		
参数	估计值	标准误
[家庭作业]截距方差	59.35	20.03
斜率方差	16.79	5.55
斜率与截距协方差	-26.27	9.68
第一层		
参数	估计值	标准误
方差 Variance	53.30	3.47
偏差度 Deviance	3 638.79	

4.5.3 NELS88完整数据库的重新分析

在本节当中,我们以完整的NELS88数据库,重新分析三个有关[家庭作业]、[生师比]、[作业×生师比]的模型。三个模型都包括了带有随机系数的[家庭作业]变量,结果全部列于表4.27以便于比较。

表4.27的上半部是固定效果,下半部为随机效果。第一个模型包括了上述三个变量,第二个模型仅有一个[生师比]主要效果,因此与本节的第一个模型(模型22)相同,第三个模型则仅有交互作用项而没有[生师比]的主要效果,因此与本节的第二个模型(模型23)相同。

表 4.27 完整 NELS88 的[生师比]与跨层级交互作用的分析结果[译者分析16]

参数	新模型		模型 22		模型 23	
	估计值	标准误	估计值	标准误	估计值	标准误
[cons]	51.43	0.70	51.52	0.61	47.90	0.19
[家庭作业]	1.52	0.18	1.48	0.05	2.00	0.16
[生师比]	-0.20	0.04	-0.21	0.03		
[作业×生师比]	-0.003	0.01			-0.03	0.008
第二层						
[截距/截距]	23.29	1.51	23.28	1.51	24.19	1.55
[家庭作业/截距]	-0.91	0.31	-0.91	0.31	-1.06	0.32
[家庭作业/家庭作业]	0.52	0.10	0.52	0.10	0.54	0.10
第一层						
[截距/截距]	71.74	0.72	71.74	0.72	71.74	0.72
偏差度 Deviance	155 679		155 679		155 706	

表 4.27 的结果显示,[家庭作业]与[作业×生师比]的高相关($r=0.89$)像在小样本分析一样会影响[家庭作业]变量的估计。当交互作用项存在于模型中时(表 4.27 的新模型与模型 24),[家庭作业]的系数的标准误变得相当大。模型 24 的[家庭作业]系数数值也变大,可能是因为[生师比]变量移除的影响,[生师比]与[家庭作业]的相关为 -0.06,与[作业×生师比]的相关为 0.31(见表 4.1)。模型 24 的跨层级交互作用此时具有统计显著性,而[生师比]并没有在模型中。偏差度的数据则显示,前面两模型的拟合情形相当,表示增加跨层级交互作用并不会改变模型拟合,在这种情况下,我们很可能会把交互作用移除,只保留[生师比]的主要效果在模型中,但是这种做法是在我们从理论层面上认为交互作用并不重要的情况下才适宜。

从三个模型的分析结果中,我们可以发现,模型中如果没有[生师比]主要效果,跨层级交互作用变量的系数具有统计显著性。此外,我们发现一个非常强的[生师比]效果,在新模型中的 z 检定值为 5.00,模型 23 的 z 检定值为 7.00。

4.6 讨 论

本章的范例分析说明了一个重要的事实，就是我们不应受限于个别回归系数的假设检验，而应重视模型拟合的检验。基本上，在科学家们的真实生活中，假设检验通常没有什么功能。研究者所关心的是数据到底能够告诉他们什么，他们想要得到从实际数据分析的结果到底有多模糊的相关信息，有时候，他们会希望得知同时考虑先验理论与观察数据两者所得出的结论为何，而这些都不是系数的假设检验可以回答的问题。

更明确地说，若我们比较 23 个学校与 1 003 个学校两个数据库所得到的分析结果（第4.5.3 节），我们很难从小样本的个别参数的假设检验的 z 检定或 t 检定结果中获得清楚明确的结论。从实际的范例中，我们发现小样本的模型分析中会出现个别参数并不显著，但是模型拟合却有改善的结果。然而在大型数据库中，即使个别参数不是很重要，或系数数值很小，但却很轻易地达到统计显著水平。从这个观点来看，偏差度是一个判断模型好坏的较佳指针，然而，在样本数很大的情况下，模型中的些微改变，都可能造成显著的变化（改善）。

就像本章各部分的示范，我们发现增加变量会影响参数估计，但是模型拟合不一定会有改变，同时参数估计的变化有时候非常剧烈，例如在最后一个部分的示范中即可发现。特别是小样本的分析，增加跨层级交互作用项会造成参数估计的不稳定性，并伴随着标准误的提高。在多层次分析中，我们所谓的小样本主要是指学校数的多寡而不是个别观察值的多寡，例如我们的小样本数据库共有 519 名学生，一般来说 $n=519$ 并不是小样本，但是在多层次分析中，23 所学校的数据库即是小样本。因此，有关小样本与全整样本的差异是要好好思考的一个问题，在本章的示范中，我们可以发现很多不同，特别是标准误与第二层的估计值，差异十分明显。

我们要再次提出呼吁，在多层次分析的实务层次中，当我们增加跨层级交互作用或脉络变量（例如取个体数据的平均值的聚合变量）时，模型有扩张的趋势。因此在进行多层次分析之前，数据的缩减是一个非常必要的程序。多层次分析并不是用来探索大量变量关系的技术，而是用来检验少量的变量及当模型每一次发生一些小小改变所产生的影响。就如同在本章的实例中可以发现，模型中的小小变化，可能会造成结果的大大改变。

批 注

1. 对于基本的(basic)变量,是指从原来数据库中被挑选出来那些变量,但不包括后来所创造的变量,例如交互作用与聚合变量(平均数)。

2. 除了第2章注2所列举的符号系统,我们必须注意在MLn软件所使用的变量的代号的差别。MLn所使用的标签最多只能有八个字符。MLn的标签是用在长方盒状的指令中,撰写时不用大写。

3. 本书所使用的数据库可另行提供给读者来进行演练,特别是MLn软件的应用。

4. 在这种情况下,截距被称为变量,即使它是个常数,亦即当所有的观察值都有相同的截距时。

5. 这表示方差成分的定义是对应于特定变量。

6. 在本书中,解释变量的组间变动仅发生在该变量被测量层次的更高层次,但有一个例外是截距cons的系数。MLn允许使用者把系数在它自己的层次或更高的层次上设定为随机,此一可以让系数在自己的层次上设为随机的这个可能性,在本书并不进行讨论,而且这个议题超过了多层次分析的范畴。

7. 两个阶层有两个R^2,三个阶层有三个R^2。这是假设误差项没有跨层级的相关。有关两个层次的两个R^2的计算、应用、意义的细节将在第5章讨论。

8. 当随机斜率存在时,组间方差会变得很复杂,这已经在稍早讨论过。

9. 固定(fixed)是相对于随机(random)而言。固定意味着我们不假设[家庭作业]对于[数学成绩]的效果在学校间具有差异。

10. 相关系数可以利用协方差(-26.14)除以斜率与截距标准差的乘积求得。

11. 多层次模型的初次使用者可能会把第二层变量与脉络变量两个概念搞混。从概念上来说,像[父母教育程度]这种变项可能可以被定义成脉络变量,但是不被视为是第二层解释变量。基本上,第二层(或脉络)的定义非常明确,在我们的例子中,第一层是学生,第二层是学校。在多层次分析中,一个变量所属的层级,最好从数据矩阵中来判断,如果一个变量在一个脉络(学校)之中数值相同,没有任何变异(同一个学校的每一个学生都有相同的数值),此变量可以被定义成第二层变量。[父母教育程度]是以学生层次变量来定义,而不是脉络变量。在我们的例子中,脉络被定义成学校,同一个学校的学生的[父母教育程度]并不相同。[父母教育程度]当然有可能被聚合成一个学校层次变量。聚合的方法有很多种,最普遍的方法是求平均数(每一个学校的父母教育程度平均值)。

12. 值得注意的是,在MLn软件中,一个系数在该变量被测量的相同层次上可以被设定为随机。做法很简单,例如我们只需把第二层的变量设定第二层方差即可(SETV 2 'schsize'),此一特别的MLn选项我们并不讨论。在本书当中,我们假设使用者仅会把一个变量的系数在比他们自己被测量层次的较高层次上设定为随机。在多层次分析中,像学校规模或学校类型这种变量,仅可能在第三阶层被设定为随机。在我们的数据库中,第三阶层可能是美国的不同州。在三个层次的数据库中,学生嵌套于学校,学校嵌套于不同的州。如果我们对于州别对学校的影响有兴趣,我们可以让学校层级变量在不同的州当中设定为具有差异。如果是三层级的分析,我们需要更多的数据。如果在我们的例子中,把23个学校依美国的50个州加以区分,是一个没有意义的作法。

译者分析

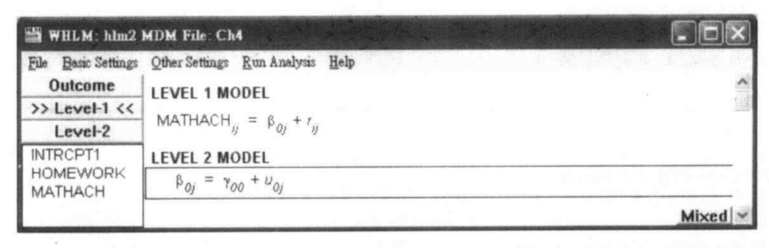
1. **模型 0:虚无模型**:以 School23_l1. sav 与 School23_l2. sav 进行 HLM6 分析,虚无模型分析的操作画面、描述统计与报表列举如下。

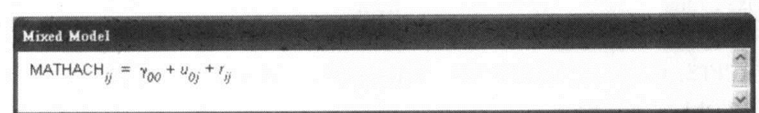
LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
HOMEWORK	519	1.97	1.48	0.00	7.00
MATHACH	519	51.72	10.71	30.00	71.00

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MEANSES	23	-0.11	0.56	-1.07	1.18
HOMEWORK	23	1.83	0.59	0.75	3.30
EDUC_MEA	23	3.15	0.90	1.95	5.00
SCHSIZE	23	3.13	1.63	1.00	7.00
WHITE	23	2.30	2.27	0.00	7.00
RATIO	23	17.35	4.88	10.00	28.00





Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	50.758 849	1.151 098	44.096	22	0.000

Final estimation of variance components:

Random Effect	Standard	Variance	df	Chi-square	P-value

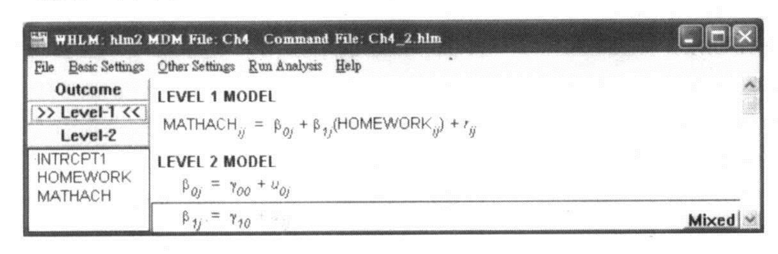
		Deviation	Component			
INTRCPT1,	U0	5.111 07	26.123 02	22	240.764 41	0.000
level-1,	R	9.013 54	81.243 93			

Statistics for current covariance components model

Deviance	= 3 798.678 611
Number of estimated parameters	= 2

上述结果与表 4.2 的数值相近,斜率与截距数值接近。偏差度为 3 798.68,与表 4.2 的 3 800.78相近。

2. **模型 1:一个微观解释变量** HomeWork:在先前的虚无模型中增加一个微观层次解释变量 HomeWork,操作画面与报表列举如下。



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	46.355 709	1.162 623	39.872	22	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	2.399 882	0.277 197	8.658	517	0.000

Final estimation of variance components:

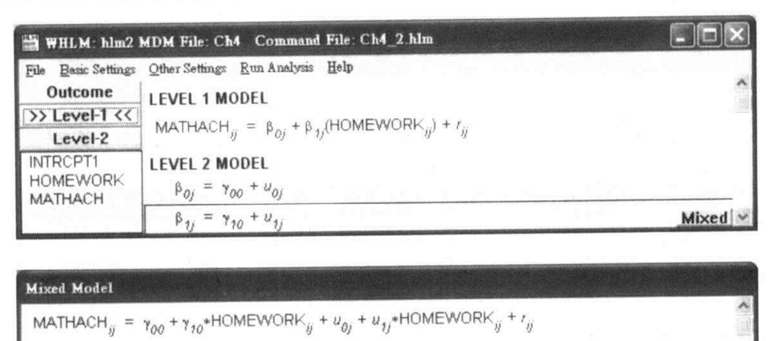
Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	4.61942	21.339 07	22	193.436 67	0.000
level-1,	R	8.442 98	71.283 91			

Statistics for current covariance components model

Deviance	= 3 727.481 647
Number of estimated parameters	= 2

上述结果与表4.3的数值相近，斜率与截距数值接近。偏差度为3 727.48，与表4.3的3 730.49相近。

3. **模型2：一个微观解释变量** HomeWork **与随机斜率**：在先前的单一解释变量模型中，令斜率为随机。操作画面与报表列举如下。



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	46.325 566	1.758 922	26.337	22	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	1.980 224	0.928 325	2.133	22	0.044

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	7.900 88	62.423 94	22	214.587 39	0.000
HOMEWORK stop, U1	4.210 08	17.724 75	22	191.518 40	0.000
level-1, R	7.300 33	53.294 76			

Statistics for current covariance components model

Deviance	= 3 633.724 203
Number of estimated parameters	= 4

上述结果与表4.4的数值相近,斜率与截距数值接近。偏差度为3 633.72,与表4.4的3 639.04相近。

4. **模型3:两个微观解释变量** HomeWork**(随机斜率)与父母教育(固定斜率)**:在先前的单解释变量模型中,HomeWork的斜率为随机系数,并增加一个解释变量EDUC。操作画面与报表列举如下。

WHLM: hlm2 MDM File: Ch4　Command File: Ch4_4.hlm

File　Basic Settings　Other Settings　Run Analysis　Help

Outcome　Level-1　>> Level-2 <<

INTRCPT2　MEANSES　PUBLIC　HOMEWORK　WHITE　RATIO

LEVEL 1 MODEL

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}(HOMEWORK_{ij}) + \beta_{2j}(EDUC_{ij}) + r_{ij}$$

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

Mixed

Mixed Model

$$MATHACH_{ij} = \gamma_{00} + \gamma_{10}*HOMEWORK_{ij} + \gamma_{20}*EDUC_{ij} + u_{0j} + u_{1j}*HOMEWORK_{ij} + r_{ij}$$

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	38.929 164	1.354 136	28.748	22	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	2.210 438	0.265 169	8.336	516	0.000
For EDUC slope, B2					
INTRCPT2, G20	2.471 685	0.320 113	7.721	516	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	3.522 64	12.408 99	22	114.227 83	0.000
level-1,	R	8.077 65	65.248 43			

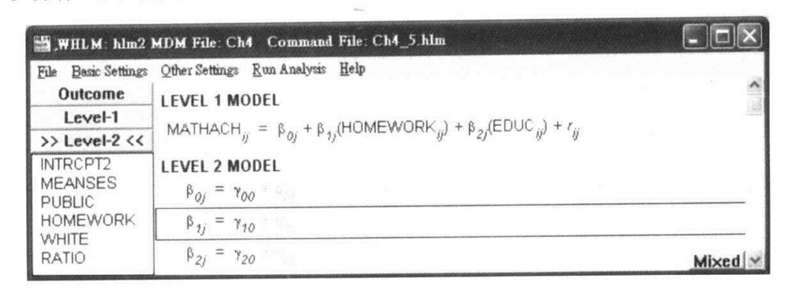
Statistics for current covariance components model

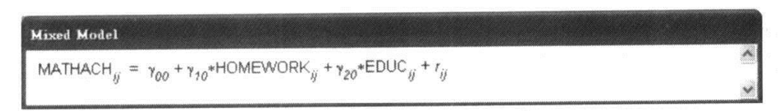
Deviance　= 3 674.876 248

Number of estimated parameters = 2

上述结果与 MLn 分析的差异较大。偏差度为 3 674.88,与表 4.5 的3 602.45尚且相近。

5. **模型 4:回归模型(HomeWork 与 EDUC 均为固定)**:在先前双解释变量模型中,完全不考虑组间差异。操作画面与报表列举如下。





Final estimation of fixed effects:

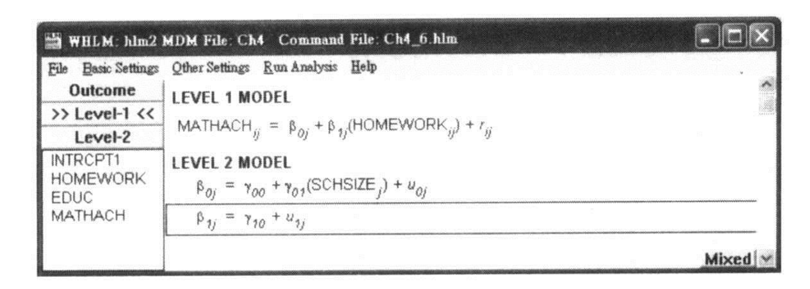
Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	37.239 150	0.996 279	37.378	616	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	2.335 361	0.268 397	8.701	516	0.000
For EDUC slope, B2					
INTRCPT2, G20	3.003 987	0.276 506	10.864	516	0.000

分析结果为传统的回归分析数据,而无随机部分的报表。固定部分的数据则与 4.2.6 节的内容相近。

6. **模型 5:完整模型(HomeWork 与 SchSize)**:微观解释变量为 HomeWork(随机系数),宏观解释变量为 Schsize。操作画面与报表列举如下。

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	44.973 309	2.710 805	16.590	21	0.000
SCHSIZE, G01	0.425 541	0.640 966	0.664	21	0.514
For HOMEWORK slope, B1					
INTRCPT2, G10	1.981 071	0.929 991	2.130	22	0.044



Mixed Model

$MATHACH_{ij} = \gamma_{00} + \gamma_{01}*SCHSIZE_j + \gamma_{10}*HOMEWORK_{ij} + u_{0j} + u_{1j}*HOMEWORK_{ij} + r_{ij}$

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	8.150 03	66.423 05	22	221.644 08	0.000
HOMEWORK slope,	U1	4.217 46	17.786 97	22	191.468 53	0.000
level-1,	R	7.301 07	53.305 60			

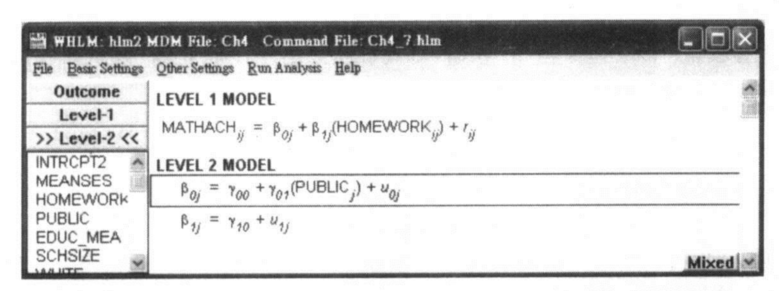
Statistics for current covariance components model

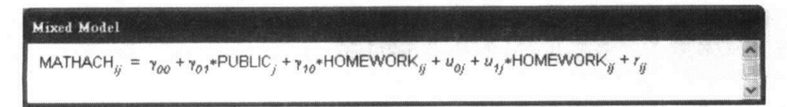
Deviance = 3 634.230 673

Number of estimated parameters = 4

分析结果与第 4.3.2 节内容一致,偏差度为 3 634.23,与表 4.7 内容相近。

7. **模型 6:完整模型**(HomeWork 与 Public):微观解释变量为 HomeWork(随机系数),宏观解释变量改为 Public。操作画面与报表列举如下。





Mixed Model

$MATHACH_{ij} = \gamma_{00} + \gamma_{01}*PUBLIC_j + \gamma_{10}*HOMEWORK_{ij} + u_{0j} + u_{1j}*HOMEWORK_{ij} + r_{ij}$

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	49.050 360	2.184 848	22.450	21	0.000
PUBLIC. G01	-4.061 008	1.977 854	-2.053	21	0.052
For HOMEWORK slope, B1					
INTRCPT2, G10	1.975 273	0.918 427	2.151	22	0.043

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	7.756 91	60.169 58	21	192.871 25	0.000
HOMEWORK slope,	U1	4.160 64	17.310 91	22	191.408 38	0.000
level-1,	R	7.303 65	53.343 35			

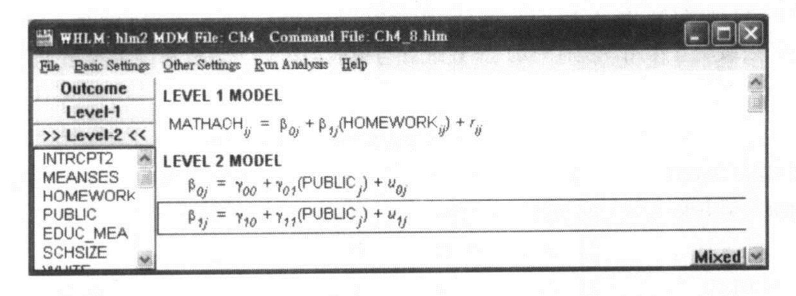
Statistics for current covariance components model

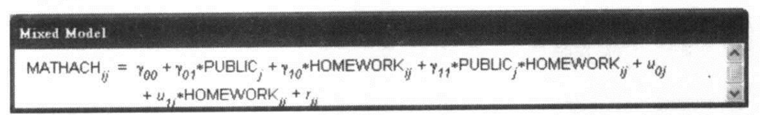
Deviance = 3 628.388 074

Number of estimated parameters = 4

分析结果与第 4.3.3 节内容一致,偏差度为 3 628.39,与表 4.8 的 3 634.84 相近。

8. **模型 7:完整模型(HomeWork 与 Public 具交互作用)**:微观解释变量为 HomeWork(随机系数),宏观解释变量改为 Public,同时这两个变量具有跨层级交互作用。操作画面与报表列举如下。





Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	48.528 899	3.016 041	16.090	21	0.000
PUBLIC, G01	-3.261 963	3.714 613	0.878	21	0.390
For HOMEWORK slope, B1					
INTRCPT2, G10	2.292 764	1.591 420	1.441	21	0.164
PUBLIC, G11	-0.495 648	1.972 838	-0.251	21	0.804

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	7.909 19	62.555 33	21	195.231 00	0.000
HOMEWORK slope,	U1	4.272 27	18.252 30	21	193.212 81	0.000
level-1,	R	7.302 92	53.332 62			

Statistics for current covariance components model

Deviance = 3 623.312 669

Number of estimated parameters = 4

分析结果与第4.3.4节内容一致,偏差度为3 623.31,与表4.10的3 634.77相近。

9. **模型8:以全部NELS88数据的完整模型(HomeWork与Public具交互作用)**:改以全体NELS数据来分析,微观解释变量为HomeWork(随机系数),宏观解释变量改为Public,同时这两个变量具有跨层级交互作用。操作画面、描述统计与报表列举如下。分析结果与第4.3.5节内容一致。

LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
HOMEWORK	21 580	1.98	1.47	0.00	7.00
SES	21 580	-0.04	0.79	-2.52	2.31
EDUC	21 580	3.14	1.28	1.00	6.00
MATHACH	21 580	51.01	10.18	26.75	71.22

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
PUBLIC	1 003	0.78	0.42	0.00	1.00
MEANSES	1 003	-0.06	0.51	-1.27	1.41
SCHSIZE	1 003	3.65	1.65	1.00	7.00
WHITE	1 003	2.92	2.20	0.00	7.00
RATIO	1 003	17.62	4.89	10.00	30.00

WHLM: hlm2 MDM File: Ch4b Command File: Ch4_9.hlm

File Basic Settings Other Settings Run Analysis Help

Outcome
Level-1
>> Level-2 <<
INTRCPT2
PUBLIC
MEANSES
SCHSIZE
WHITE
RATIO

LEVEL 1 MODEL

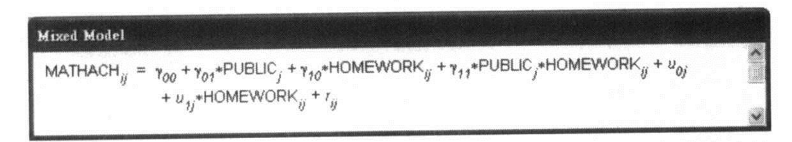
$MATHACH_{ij} = \beta_{0j} + \beta_{1j}(HOMEWORK_{ij}) + r_{ij}$

LEVEL 2 MODEL

$\beta_{0j} = \gamma_{00} + \gamma_{01}(PUBLIC_j) + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}(PUBLIC_j) + u_{1j}$

Mixed



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	52.719 676	0.380 313	138.622	1 001	0.000
PUBLIC, G01	-6.045 854	0.425 179	-14.220	1 001	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	0.915 709	0.102 951	8.895	1 001	0.000
PUBLIC, G11	0.684 227	0.116 472	5.875	1 001	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	4.271 47	18.245 49	995	2 679.217 48	0.000
HOMEWORK slope,	U1	0.663 45	0.440 16	995	1 281.825 79	0.000
level-1,	R	8.469 38	71.730 42			

Statistics for current covariance components model

Deviance = 155 540.389 679

Number of estimated parameters = 4

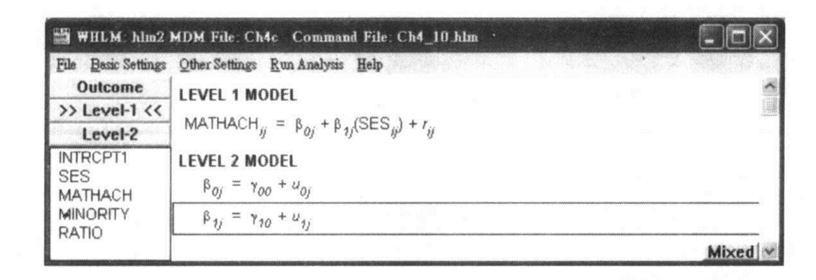
10. **模型 16:以 SES 为解释变量(固定斜率)**:以 School23_l1.sav 与 School23_l2.sav 的数据进行分析,微观解释变量为 SES,截距为随机,但是斜率为固定,因此为一个 ANCOVA 模型。操作画面、描述统计与报表列举如下。结果可以和第4.4.1节与表 4.19 的数据相比较。

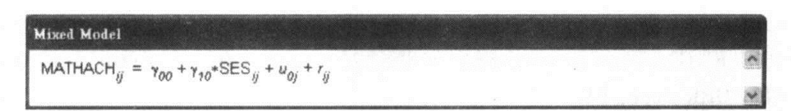
LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
SES	519	-0.00	0.88	-2.41	1.85
MATHACH	519	51.72	10.71	30.00	71.00
RATIO	519	16.76	4.93	10.00	28.00

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MEANSES	23	-0.11	0.56	-1.07	1.18
MINORITY	23	2.30	2.27	0.00	7.00
RATIO	23	17.35	4.88	10.00	28.00





Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	51.200 859	0.850 662	60.189	22	0.000
For SES slope, B1					
INTRCPT2, G10	4.332 382	0.566 265	7.651	517	0.000

Final estimation of variance components:

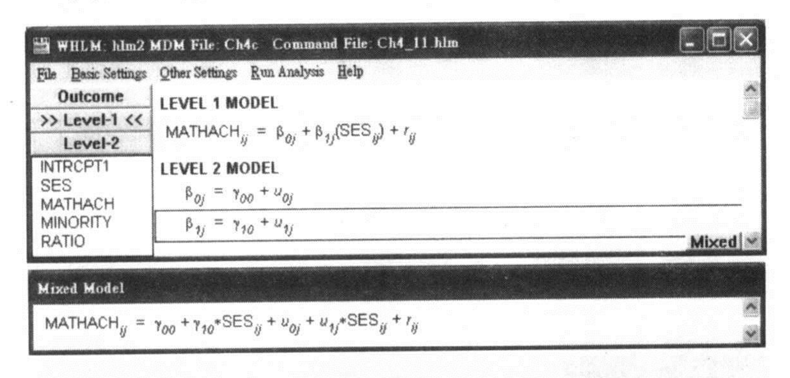
Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	3.553 94	12.630 48	22	119.556 40	0.000
level-1,	R	8.679 17	75.327 95			

Statistics for current covariance components model

Deviance = 3 744.339 554

Number of estimated parameters = 2

11. **模型 18:以 SES 为解释变量(随机斜率)**:前述的模型增加斜率为随机之设定,亦即各校的 SES 解释力不同,亦即组内回归系数不同质下的检验。操作画面与报表列举如下。结果可以与第 4.4.2 节表 4.20 的数据相比较。



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	51.249 012	0.857 867	59.740	22	0.000
For SES slope, B1					
INTRCPT2, G10	4.326 502	0.571 343	7.573	22	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	3.595 97	12.931 01	22	93.956 29	0.000
SES slope,	U1	0.409 43	0.167 63	22	29.954 93	0.119
level-1,	R	8.669 87	75.166 57			

Statistics for current covariance components model

Deviance = 3 744.195 212
Number of estimated parameters = 4

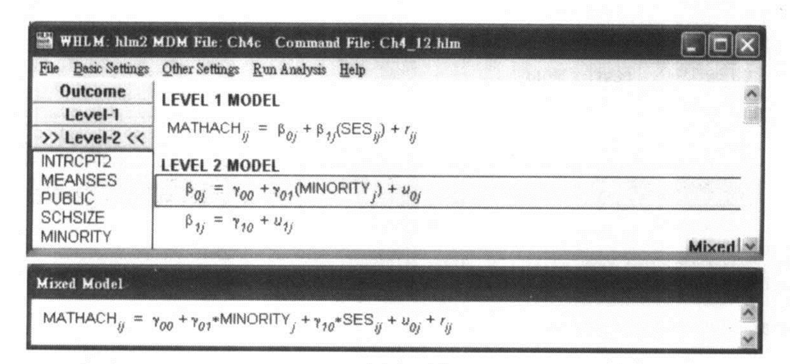
值得一提的是,表 4.20 指出第二阶层的斜率方差 MLn 因为没有收敛而无法估计,但以 HLM6 分析时,经过 581 次迭代后完成收敛,得到数据为0.167 63。

此外,斜率与截距协方差亦可估计,数值为 -1.068 45。如下所示:

Tau

INTRCPT1,B0	12.931 01	-1.068 45
SES,B1	-1.068 45	0.167 63

12. **模型 18:完整模型(SES 与 Minority)**:在固定 SES 变量模型中,增加一个 Minority 来解释数学成绩。操作画面与报表列举如下。结果可以和第 4.4.3 节与表 4.21 的数据相比较,数值相当接近。



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	53.126 616	1.177 846	45.105	21	0.000
MINORITY, G01	-0.809 365	0.364 656	-2.220	21	0.037
For SES slope, B1					
INTRCPT2, G10	4.298 969	0.561 807	7.652	516	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	3.270 43	10.695 73	21	110.569 39	0.000
level-1,	R	8.670 22	75.172 75			

Statistics for current covariance components model

Deviance = 3 741.610 872

Number of estimated parameters = 2

13. **模型 19:完整模型**(SES、MEANSES 与 Minority):将前述模型增加 MEANSES 来解释数学成绩。操作画面与报表列举如下。结果可以和第 4.4.4 节与表 4.22 的数据相比较,数值接近。

WHLM: hlm2 MDM File: Ch4c Command File: Ch4_13.hlm

File Basic Settings Other Settings Run Analysis Help

Outcome
Level-1
>> Level-2 <<
INTRCPT2
MEANSES
PUBLIC
SCHSIZE
MINORITY

LEVEL 1 MODEL

$MATHACH_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij}) + r_{ij}$

LEVEL 2 MODEL

$\beta_{0j} = \gamma_{00} + \gamma_{01}(MEANSES_j) + \gamma_{02}(MINORITY_j) + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

Mixed

Mixed Model

$MATHACH_{ij} = \gamma_{00} + \gamma_{01}*MEANSES_j + \gamma_{02}*MINORITY_j + \gamma_{10}*SES_{ij} + u_{0j} + r_{ij}$

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	53.096 904	1.102 667	48.153	20	0.000
MEANSES, G01	2.805 075	1.479 369	1.896	20	0.072
MINORITY, G02	−0.692 189	0.345 691	−2.002	20	0.059
For SES slope, B1					
INTRCPT2, G10	3.884 760	0.669 773	6.371	515	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	2.974 35	8.846 75	20	77.372 84	0.000
level-1,	R	8.674 13	75.240 51			

Statistics for current covariance components model

Deviance = 3 733.777 145

Number of estimated parameters = 2

14. **模型 20 与 21:以全部 NELS88 数据的完整模型(SES 与 MINORITY)**:改以全体 NELS 数据来分析微观解释变量为 SES 的随机斜率(模型 17)与固定斜率(模型 18)两个模型,宏观解释变量为 MINORITY。操作画面、描述统计与报表列举如下。分析结果与第 4.3.5 节内容一致。但值得一提的是,模型 17 需要高达 581 迭代才完成收敛,但是以 NELS88 全体数据仅需 36 个迭代就完成模型 20 的估计,显示大样本的稳定度高。

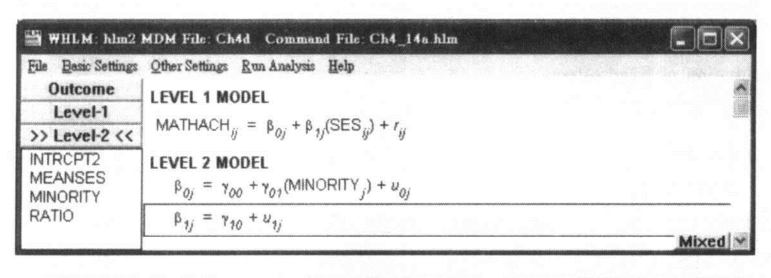
LEVEL-1 DESCRIPTIVE STATISTICS

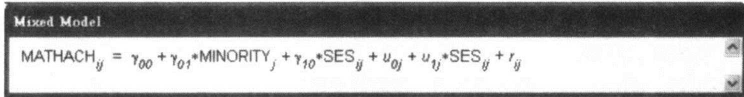
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
SES	21 580	-0.04	0.79	-2.52	2.31
MATHACH	21 580	51.01	10.18	26.75	71.22

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MINORITY	1 003	2.92	2.20	0.00	7.00

模型 20(随机斜率)





Tau

INTRCPT1,B0	8.618 99	1.554 14
SES,B1	1.554 14	0.977 39

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	53.019 885	0.183 085	289.592	1 001	0.000
MINORITY, G01	-0.701 021	0.049 459	-14.174	1 001	0.000
For SES slope, B1					
INTRCPT2, G10	4.722 316	0.093 508	50.502	1 002	0.000

Final estimation of variance components:

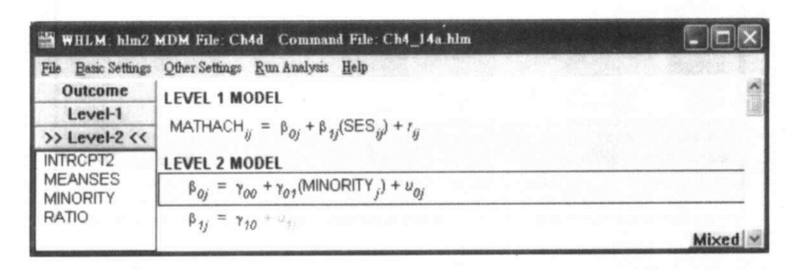
Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	2.935 81	8.618 99	997	2 941.029 07	0.000
SES slope,	U1	0.988 63	0.977 39	998	1 265.307 08	0.000
level-1,	R	8.343 12	69.607 66			

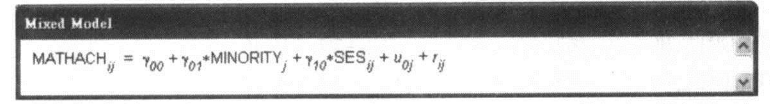
Statistics for current covariance components model

Deviance = 154 161.398 591

Number of estimated parameters = 4

模型 21(固定斜率)





Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	53.159 701	0.186 969	284.323	1 001	0.000
MINORITY, G01	-0.710 193	0.051 580	-13.769	1 001	0.000

For SES slope, B1					
INTRCPT2, G10	4.762 326	0.087 482	54.438	21 577	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	2.967 66	8.807 03	1001	3 732.920 52	0.000
level-1,	R	8.362 66	69.934 03			

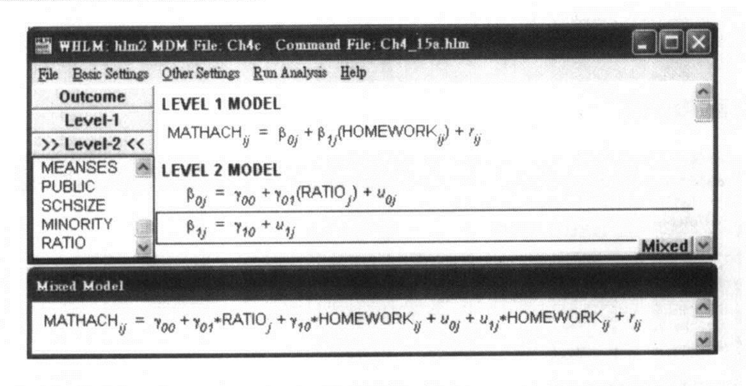
Statistics for current covariance components model

Deviance = 154 192.380 873

Number of estimated parameters = 2

15. **模型 22 与 23(跨层级交互作用有无)**:以 HomeWork(微观层次)与 Ratio(宏观层次)来解释数学成绩,若没有跨层级交互作用为模型 22,有跨层级交互作用但是移除 Ratio 主要效果的模型为模型 23,操作画面与报表列举如下。分析结果和第 4.5.1 与 4.5.2 节内容一致。

模型 22(无交互作用的完整模型)



Tau

INTRCPT1,B0	63.479 67	-27.703 46
HOMEWORK,B1	-27.703 46	17.741 87

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	47.961 614	4.083 201	11.746	21	0.000
RATIO, G01	-0.094 423	0.212 384	-0.445	21	0.661
For HOMEWORK slope, B1					
INTRCPT2, G10	1.979 281	0.928 857	2.131	22	0.044

Final estimation of variance components:

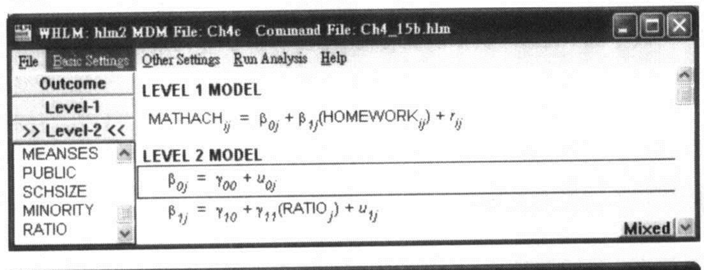
Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	7.967 41	63.479 67	21	211.414 31	0.000
HOMEWORK slope,	U1	4.212 11	17.741 87	22	191.496 32	0.000
level-1,	R	7.300 98	53.304 31			

Statistics for current covariance components model

Deviance = 3 636.637 671

Number of estimated parameters = 4

模型23(有交互作用,但是没有 RATIO 主效果的模型)



Tau

INTRCPT1, B0	62.554 51	-27.799 39
HOMEWORK, B1	-27.799 39	18.094 03

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	46.330 611	1.760 574	26.316	22	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	2.884 107	2.156 618	1.337	21	0.196
RATIO, G11	−0.052 584	0.112 271	−0.468	21	0.644

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	7.909 14	62.554 51	22	214.546 64	0.000
HOMEWORK slope,	U1	4.253 71	18.094 03	21	194.544 22	0.000
level-1,	R	7.300 54	53.297 82			

Statistics for current covariance components model

Deviance = 3 637.890 963

Number of estimated parameters = 4

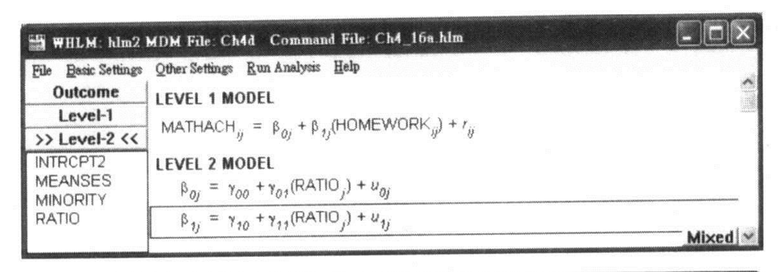
由这些报表对照表 4.25 与 4.26 的数据,发现对标准误的估计,乃至于方差的估计有些微差异,可能是取样的差异、MLn 与 HLM 估计设定的差异所造成,但是不至于影响结果的解释。

16. **模型 22 与 23:以全部 NELS88 数据的完整模型(跨层级交互作用)**:前述模型改以全体 NELS 数据来分析,共有三个模型:模型 A 有微观解释变量 HomeWork 与宏观解释变量 Ratio,以及跨层级交互作用。模型 B 则没有交互作用(模型 22),模型 C 有交互作用,但是移除 Ratio 的主要效果(模型 23),操作画面与报表列举如下。分析结果与第 4.5.3 节内容一致。

模型 A(具交互作用的完整模型)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	51.428 679	0.699 058	73.569	1 001	0.000
RATIO, G11	−0.199 725	0.038 113	−5.240	1 001	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	1.522 710	0.182 348	8.351	1 001	0.000
RATIO, G01	−0.002 737	0.010 003	−0.274	1 001	0.784



Mixed Model

MATHACH$_{ij}$ = γ_{00} + γ_{01}*RATIO$_j$ + γ_{10}*HOMEWORK$_{ij}$ + γ_{11}*RATIO$_j$*HOMEWORK$_{ij}$ + u_{0j} + u_{1j}*HOMEWORK$_{ij}$ + r_{ij}

Final estimation of variance components:

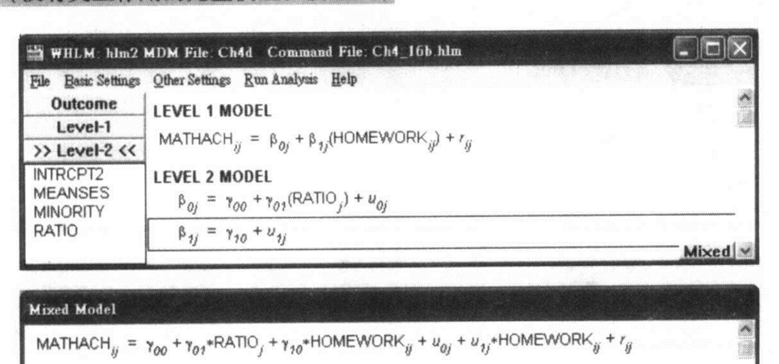
Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	4.832 62	23.354 20	995	3 140.686 16	0.000
HOMEWORK slope,	U1	0.723 01	0.522 75	995	1 339.084 15	0.000
level-1,	R	8.469 81	71.737 64			

Statistics for current covariance components model

Deviance = 155 697.233 748

Number of estimated parameters = 4

模型 B(没有交互作用的完整模型。模型 22)



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	51.522 033	0.609 025	84.598	1 001	0.000
RATIO, G01	-0.204 991	0.032 803	-6.249	1 001	0.000
For HOMEWORK slope, B1					
INTRCPT2, G10	1.474 319	0.049 112	30.019	1 002	0.000

Statistics for current covariance components model

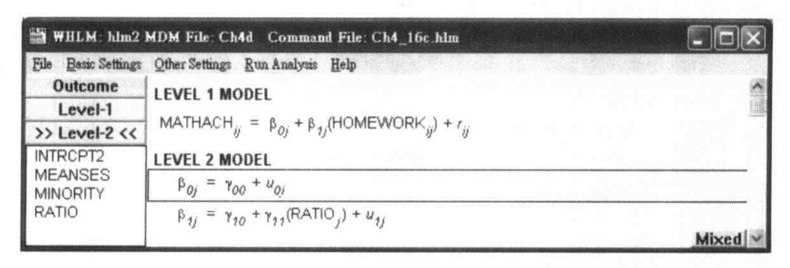
Deviance = 155 688.100 163

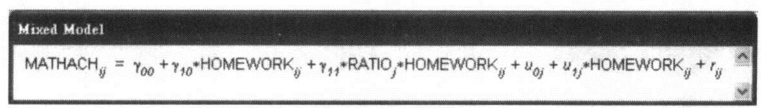
Number of estimated parameters = 4

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	4.830 44	23.333 11	995	3 140.833 50	0.000
HOMEWORK slope,	U1	0.724 19	0.524 45	996	1 337.264 55	0.000
level-1,	R	8.469 47	71.731 89			

模型 C(有交互作用,无 Ratio 主要效果的完整模型。模型 23)





Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d. f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	47.897 513	0.187 282	255.750	1 002	0.000

For HOMEWORK slope, B1					
INTRCPT2, G10	2.005 700	0.159 072	12.609	1 001	0.000
RATIO, G11	-0.030 022	0.008 632	-3.478	1 001	0.001

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	4.925 01	24.255 76	996	3 235.569 95	0.000
HOMEWORK slope,	U1	0.746 47	0.557 22	995	1 359.580 85	0.000
level-1,	R	8.469 27	71.728 59			

Statistics for current covariance components model

Deviance = 155 717.828 679
Number of estimated parameters = 4

A,B,C 三个模型的检验结果与表 4.27 几乎完全相同,显示大样本分析结果的稳定性高。有关 HLM6 的操作方法,读者可以参考软件的指导手册,或到网站 http://www.ssicentral.com/ 下载相关文件。

多层次分析的重要议题 5

5.1 绪 论

虽然多层次模型是一种相当受到限制的模型，但是却是很重要、具有高度类化性的一种从线性回归模型所衍生出来的统计方法。换言之，有关回归系数的稳定性、多元共线性的影响、模型假设的检验等这些在回归模型中就已经是困扰的大问题，在多层次模型中势必更加的复杂。而其他一些比较小的问题，在面对多层次模型的特殊构造之时，也必须有一个新的解答或做必要的修正。例如解释变量的中心化（平减）[1]，当我们有多重变异来源所造成的估计问题，等等。正因为多层次模型具有阶层性的结构，具有特殊的跨层级的交互作用，因此对于解释变量的中心化处理也就特别的重要。而面对不同层级的多重变异来源这种特殊现象，使得我们在定义解释变异（explained variance）与模式变异（modeled variance）时显得格外的复杂，而且要以不同以往的方法来界定。但是，如果是只具有单一层次的模型，或是脉络模型，问题就不会这么复杂了。

为了顺应因特网的快速发展，本章我们以逐点式的常见问题（Frequently Asked Questions；FAQ）的形式来进行几个重要议题的讨论。本章所汇集的 FAQ 问题，多是从多层次邮件社群中的讨论所延伸而来（注 1）。从这些 FAQ 问题中，我们可以发现使用多层次分析的研究者多会关心哪些问题。同时由于多层次分析软件越来越多，研究者现在可以处理越来越复杂的模型，对于结果分析与解释的问题也越来越多。以下我们就分别讨论六个常见的问题：

- 将解释变量以总平均数进行中心化（centering）（总平减）及以组平均数进行

---

[1] 本书将中心化译为平减，系取“减去平均数”之意，与经济学上的术语“平减指数”（implicit price deflator）（代表物价的变动）无关。平减若以总平均数为之，称为总平减，以组平均数为之则称为组平减。平减只是减去平均数，亦即以平均数进行平移，并非标准化（因为并没有除以标准差）。

中心化(组平减)等数学转换的影响为何?

- 解释变异量(explained variance)在阶层线性模型中是如何定义的? 阶层线性模型是不是有与 R^2类似的概念? 我们可不可以说结果变量的百分之多少的变异归因于团体因子,多少百分比的变异归因于个体因子? 如果我们增加或减少变量,所估计的统计数据会有如何改变?
- 如果使用阶层线性模型,那么我们应如何交代回归分析的统计检定力(power)?
- 多层次分析的系数可以是非随机且对于各组都是固定值,也可以是非随机但是在各组间存在着变动的数值,也可以是随机的(也就是自成一个变量)。这些不同的做法应如何选择? 不同的选择之间有何差异?
- 什么是 FIML, REML, EM, IGLS, RIGLS, EB/ML, OLS, GLS?
- 多元共线性(multicollinearity)在多层次分析当中有多么严重?

与前面的章节相比,本章无可避免地涉及较多的方法学与技术层次的讨论。但是我们会强调概念的理解,说明中尽量用非正式的数学语言,同时以前面范例及结果来解释说明。

5.2 中心化的影响

常见问题一:解释变量以总平均数或组平均数进行中心化的影响为何?

中心化的影响,我们将以第四章当中的 NELS88 数据库子样本的数据分析来说明。基本上,中心化问题并不是一个可以简单回答的问题,因为中心化的影响会随着数据型态、变项关系与研究目的不同而有所不同。读者可能会认为,多层次分析的数据操作就好像传统回归分析一样具有类似的效果,对于解释变量加减一个常数并不影响数据的关系。也正因为如此,对于中心化的影响,我们将以传统分析的观点来介绍,然后进一步地去说明在多层次回归中中心化影响的差异,以及所造成的特殊且无法预期的作用。

5.2.1 固定效果回归模型的中心化

Aiken 与 West(1991)曾经把传统回归模型的中心化问题进行完整深入的讨论。中心化是指把每一个个体的解释变量都减去某一个数值,通常是总平均数,但也可能是其他数值,来将变量的中心点进行位移。基本上,每一个分数减去同一个数值并不会改变数据间的关系。Aiken 与 West(1991)指出,如果模型中仅有第一阶层的变量,将一个变量进行加减对于变量的方差不会有影响,乃至于与其他变项的协方差与相关也都不会有任何改变。在传统的固定效果回归分析中,截距就好

像是一个自由估计的参数，使得模型的稳定性得以确保，让解释变量的原点得以维持在某一个特定的位置。中心化的作用，是使分数改成离差分数形式（deviation form），但是不影响回归系数的强度，仅影响截距的数值。换句话说，中心化改变了截距，但是回归系数维持恒定（注 2）。

实际上，传统固定效果线性模型之所以要进行中心化，目的在于改变截距的意义以便于解释。以离差分数来代替原始分数影响截距，也就影响了它的解释与应用方式。以原始分数所建立的模型，截距是指当所有的解释变量数值为 0 时的反应变量的数值。在社会科学领域中，以回归分析来探讨的研究问题，例如态度或智力测验对学习的影响，解释变量并不会有 0 的数值，0 并没有任何的意义。在这种情况下，解释变量的中心化使得截距具有意义，如果以平均数作为新的中心点，截距所反应的是解释变量为平均数时的反应变量的数值。

5.2.2　多层次模型的中心化

为了从技术层面来回答多层次模型中心化的影响这个问题，我们引用 Kreft 等人（1995）的著作，但是在说明上做了些微的改变。首先，如果一个模型在第一阶层有一个解释变量 x，带有一个随机的截距、一个随机的斜率，此时，第二阶层的解释变量 z 可能会与截距和斜率产生交互作用，得到下列多层次模型：

$$\underline{y}_{ij} = \underline{\alpha}_j + \underline{\beta}_j x_{ij} + \underline{\varepsilon}_{ij} \tag{5.1a}$$

$$\underline{\alpha}_j = \alpha + \gamma_0 z_j + \underline{\delta}_{0j} \tag{5.1b}$$

$$\underline{\beta}_j = \beta + \gamma_1 z_j + \underline{\delta}_{1j} \tag{5.1c}$$

进行整理合并后，三个方程式可整合为一个混合模型方程式。方程式的固定部分如下：

$$\mathrm{E}(\underline{y}_{ij}) = \alpha + \beta x_{ij} + \gamma_0 z_j + \gamma_1 x_{ij} z_j \tag{5.2a}$$

随机部分如下：

$$\underline{y}_{ij} - \mathrm{E}(\underline{y}_{ij}) = \underline{\delta}_{0j} + x_{ij}\underline{\delta}_{1j} + \varepsilon_{ij} \tag{5.2b}$$

我们所关心的是学生层次的解释变量 x。公式 5.1 中的 x 变量是以原始分数形式存在，如果进行总平均中心化（总平减），x_{ij} 即可用 $\ddot{x}_{ij}$ 来代替，该变量是取距总平均值的离差 $\ddot{x}_{ij} = x_{ij} - \bar{x}$。由于每一个人所减掉的数值为同一个常数（所有的人都减掉同一个平均数），因此用此中心化分数（总平减分数）所得到的结果，除了截距 α 不同之外，其他都与使用原始分数得到的结果一样。

总平减比组平减（以组平均数来进行中心化）简单许多。由于各组的组平均并不相同，每一组（脉络）的个体所减掉的数值并不相同，组平减后的解释变量 $\bar{x}_{ij}$ 为

原始分数减去各组平均数 $\tilde{x}_{ij} = x_{ij} - \bar{x}_j$ 。使用此一组平减变量来取代原始分数形式变量，将使模型产生变化。若将组平减变量带回方程式后重建等式，组平均数在方程式的角色就如同其他第二层次的解释变量，如下式所示：

$$\underline{\alpha}_j = \alpha + \gamma_{01}\bar{x}_j + \gamma_{02}z_j + \underline{\delta}_{0j} \quad (5.3)$$

方程式中增加了平均数之后，对于其他参数估计将产生影响，我们将以实际范例来说明。

当随机斜率加入之后，统计的恒等性将不复存在，但是值得注意的是，统计的恒等性即使存在，参数估计值也未必相同。从实际运用的角度来看，组平均数的中心化与原始分数所建立的模型是不相同的，在此必须强调的是，组平均数的中心化有其不同的使用目的，可用于检验不同的理论。

5.2.3 总平均中心化(总平减)

我们先来讨论用解释变量进行总平减的影响。首先，解释变量可拆解成平减后总和与总平均数两个部分。若把公式 5.2a 与 5.2b 中的解释变量 x_{ij} 改用离均差 $\ddot{x}_{ij}$ 与总平均 $\bar{x}$ 来表示，然后将每一项依照下标加以整理，我们会发现两个固定系数即截距 α 的数值与第二阶层的系数 γ_0 产生了改变。总平减后的方程式元素 $\beta\bar{x}$ 与 $\gamma_1\bar{x}$ ，分别加入了截距 α 与第二层斜率 γ_0 ，如下所示：

$$\mathrm{E}(\underline{y}_{ij}) = [\alpha + \beta\bar{x}] + \beta\ddot{x}_{ij} + [\gamma_0 + \gamma_1\bar{x}]z_j + \gamma_1\ddot{x}_{ij}z_j \quad (5.4a)$$

公式 5.4a 当中，两个括号内都增加了一项被总平均数加权的系数。用总平减变量的系数可以轻易地计算出以原始分数来进行估计的系数值。当一个模型使用总平减变量来进行参数估计时，回归方程式当中的两个中心化前的系数(以原始分数来计算得出的系数)可以直接将 α 或 γ_0 减去平均数乘以 β 或 γ_1 来得出。

另一个可以明显看到模型产生改变之处，是截距方差的数值。方差变化量仍与平均数有关。以公式 5.4a 进行类似的整理后，可以得到下式：

$$\underline{y}_{ij} - \mathrm{E}(\underline{y}_{ij}) = (\underline{\delta}_{0j} + \bar{x}\,\underline{\delta}_{1j}) + \ddot{x}_{ij}\underline{\delta}_{1j} + \varepsilon_{ij} \quad (5.4b)$$

很明显，使用原始数据与总平减数据进行估计，虽然参数估计值会发生变动，但是并没有改变模型的组成与变量的关系。将公式 5.2a 与公式 5.4a、公式 5.2b 与公式 5.4b 加以比对，基本上都是相同的模型，前者使用原始的解释变量 x_{ij} ，后者使用平减后的解释变量 $\ddot{x}_{ij}$ 。依照 Kreft 等人(1995)所使用的术语，原始分数模型与总平减模型为等价线性模型(equivalent linear model)。但这并不是说所有的参数估计结果是真的相等，等价模型会有相同的拟合度，相同的预测值，相同的残差。至于参数估计的数值，则可利用数学方法来证明其间的关系。

5.2.4 分组平均中心化(组平减)

接下来,让我们看看如果把解释变量 x_{ij} 用分组平均数进行组平减会有什么影响。此时组平减变量为 $\tilde{x}_{ij} = x_{ij} - \bar{x}_j$,而 $\bar{x}_j$ 为组平均数。如同第5.2.3节的程序,我们可以把 $x_{ij} = \tilde{x}_{ij} + \bar{x}_j$ 代入公式5.5中,再把各项加以整理,会得到与前面总平减类似的参数组合的变化,只是我们要把公式5.4a当中的总平均数改成组平均数。也就是:

$$E(\underline{y}_{ij}) = [\alpha + \beta\bar{x}_j] + \beta\tilde{x}_{ij} + [\gamma_0 + \gamma_1\bar{x}_j]z_j + \gamma_1\tilde{x}_{ij}z_j \quad (5.5a)$$

对于随机部分,则为:

$$\underline{y}_{ij} - E(\underline{y}_{ij}) = (\underline{\delta}_{0j} + \bar{x}_j\,\underline{\delta}_{1j}) + \tilde{x}_{ij}\underline{\delta}_{1j} + \varepsilon_{ij} \quad (5.5b)$$

公式5.5a再一次显示了截距与第二层解释变量 z_j 的系数受到了影响,但是此时问题出在我们不再可能将这两项还原成原始分数的数据型态。公式5.5a中的平均数不像先前的总平均数为单一数值,而是每一组都不同,导致每一组的截距 α 被减去一个不同的数值,因而没有一个共同的截距值。相同的情况发生在公式5.5a的第二层解释变量 z 的系数和公式5.5b的截距方差。

组平减模型与原始分数模型的相异不仅在固定部分,也在随机部分。但是在两种情况下则不然,第一是每一组的平均数 $\bar{x}_j$ 等于总平均数 $\bar{x}$,此种状况可能发生在重复量数研究中,但几乎不会发生在社会科学与组间差异有关的研究中(例如学校效能研究)。

另外一种会发生组平减与原始分数两种模型为等值模型的情况,是当模型中只有随机截距,而斜率为常数(非随机)时,此时组平均数又变成了第二层的解释变量。简单来说,当 $\gamma_1 = 0$ 时,斜率为固定值,因此 $\underline{\delta}_{1j} = 0$,且 $z = \bar{x}_j$。此时,模型的拟合与仅带有一个随机截距、但第二层次没有平均数为解释变量的原始分数模型的拟合会相等,这两种简化的模型为等值模型,组平减模型的参数估计值可以通过数学过程转换出与这种简化原始分数模型相同的数据。

5.2.5 范例说明

有许多文献指出,私立学校对于学生的学习成绩有正向的效果(Coleman et al., 1982),其中一个原因是私立学校较公立学校具有平等主义(egalitarian)的倾向。以下的章节,我们利用MELS88的数据来检验这个现象,同时通过平减模型与原始分数模型的比较来说明研究结果可能出现的差异。范例中,解释变量的处理方式分别有脉络内的平减(centered within contexts; CWC)与原始分数(raw scores; RS)两种不同做法。

Raudenbush 与 Bryk(1986)对高中及以上层级学校的数据库所进行的研究结果表明,私立学校在[SES]与[数学成绩]两变量之间具有显著的关系,亦即学校层次的解释变量[公立]与学生层次解释变量[SES]对于[数学成绩]的交互影响具有统计的显著性,此即为跨层级交互作用。Raudenbush 与 Bryk(1986)的研究数据表明,天主教学校较具有公平主义的精神,在天主教学校中,"较低社经地位的学生在天主教学校表现得比较好,但是高社经地位的学生在公立学校表现得比较好"(Raudenbush & Bryk, 1986, p. 13)。我们重复了 Raudenbush 与 Bryk(1986)的模型,并以 NELS88 的数据库中的21 580 位学生与1 003 所学校的数据来检验私立与公立学校的差异,所使用的模型如同前面几章一样,以[数学成绩]为反应变量,用学生层次解释变量[家庭作业]与[SES]来进行预测。在 NELS88 数据库中,大多数学校为公立学校(80%),其余 20% 可以区分为天主教学校(10%)、宗教学校(4%)与非宗教学校(6%)。在 Raudenbush 与 Bryk 的模型中,变项均经过了平减(CWC)处理,而我们则是同时以 CWC 与 RS 两种模式来进行分析。

分析过程中,我们提出了三种模型来匹配观察数据,所使用的解释变量都是[家庭作业]与[SES],但是平减程序则有不同。各模型的第二层解释变量都有[公立],但有的模型则另外增加了[生师比]。三个模型分别为 RS 模型、CWC 模型但没有将平均数置回第二层作为解释变量(CWC(N))、CWC 模型但把平均数置回第二层作为解释变量(CWC(M))。

这些模型有时候会被误以为是等值模型,甚至是没有把平均数置回第二层作为解释变量的这种模型也会以为与其他两种模型相等。在第三种模型(CWC(M))中,变量经过平减,但被减去的平均数则被置回第二层作为解释变量,此时平减变量的组间变异被移除了,但是平均数的差异在第二层获得还原。被移去的是解释变量的组间变异,此一组间变异在整个模型中扮演重要的角色。在我们的实际范例中即可看出,以平均数进行中心化平减,造成变量在各组间的变异被移除,包括[家庭作业]与[SES]两个变量,对于公私立之别变量的效果影响很大(见表5.1 与 5.2)。这两个表的差别在于模型的不同,包括 RS,CWC(N),CWC(M)。最后一个模型是 Raudenbush 与 Bryk 所使用的模型。

表 5.1 中,平减后的变量以加底线来表示,如[<u>SES</u>]与[<u>家庭作业</u>]。[家庭作业]在 RS 模型中被使用,而[<u>家庭作业</u>]被使用在两个 CWC 模型中。所有的模型都有[公立]这个变量,私立学校编码为 0,公立学校编码为 1。值得注意的是表5.1的第三个模型多了两个第二层次解释变量[SES 平均数]与[家庭作业平均数]。在 RS 模型,平均数并没有纳入模型中,因为并没有任何的平减程序,因此,平均数的信息仍保留在模型中。表 5.1 的模型对于[SES]与[家庭作业]都是设定为随机斜率。

表 5.1　各变量在不同模型的效果(加底线的变量是经过平减者)

	RS		CWC(N)		CWC(M)	
	EST	SE	EST	SE	EST	SE
截距	50.16	0.28	55.06	0.35	47.53	0.47
[家庭作业]	1.24	0.05	N/A		N/A	
[家庭作业]	N/A		1.18	0.05	1.20	0.05
[SES]	4.35	0.09	N/A		N/A	
[SES]	N/A		3.84	0.10	3.85	0.10
[公立]	-2.06	0.29	-5.42	0.39	+0.62	0.28
[SES 平均数]	N/A		N/A		8.14	0.25
[家庭作业平均数]	N/A		N/A		1.65	0.20
[SES]斜率方差	0.47	0.33	1.65	0.44	1.71	0.44
偏差度 Deviance	153 333		153 986		153 004	

如表 5.1 所示,不管在系数部分还是在模型拟合度(偏差度)上,RC 模型与其他两个模型的结果也不相同,而两个 CWC 模型也不同。差异的主因是 CWC(N)模型对于[SES]与[家庭作业]两个变量的学校间差异被移除的结果。平均数的差异经平减后从原始资料中被去除且没有被置回。按理说,CWC(M)模型应该与 RS 模型最接近,CWC(N)就不然。从数据中可知的确是 CWC(N)的拟合度最差(偏差度最高)。

固定效果的差异主要是在[公立]变量的系数与截距。[家庭作业]与[SES]系数的估计值,在原始与平减模型中都非常接近,标准误也非常接近相等。在随机部分,我们仅列出了[SES]的斜率方差,而 RS 与 CWC 模型间的差异得到了相反结论。三个模型对于[SES]的斜率方差的估计值数值都不同,但标准误则非常接近。在 RS 模型中,方差并不显著,但是在两个 CWC 模型中都是显著的。

相同的差异现象也发生在[公立]的固定效果上。在 RS 模型中,[公立]变量具有高度的负向效果,系数为 -2.06,在 CWC(N)模型强度更高,系数达 -5.42,但是在 CWC(M)模型中则改变了方向,成为 +0.62,数值虽小但是达到显著水平。这三个系数均具有统计的显著性,但强度不同,更重要的是方向也不同。很明显,公私立之别的效果可以从负向效果变成正向效果,这取决于我们如何处理原始数据。平减但不置回平均数对于私立学校最有利,也就是 CWC(N)模型的结果,从原始的 RS 模型的 -2.06 提高到 -5.42。平减但置回平均数对于公立学校有利,[公立]的效果量变成 +0.62,会变成对于公立学校有利的结果是缘于平均数被纳入估计,也就是说,[SES]与[家庭作业]的校间差异被还原,造成结论的改变。

表 5.1 的三个模型并不是等值模型,可以从偏差度指标中看出,这三个模型对

于数据的拟合并不相等,CWC(N)模型的偏差度指针最高,也就是对于观察数据有最差的拟合,这并不令人惊讶,因为CWC(N)模型已把[SES]与[家庭作业]的组间差异移除,平均数亦未置回,改变幅度最大。另外,随机部分有明显差异,是在RS模型中,[SES]斜率的随机性并没有达到统计的显著性。但是两个CWC模型则达到统计的显著性。这些发现会影响后续的分析,我们将在下面的章节中继续介绍。

5.2.6 跨层级交互作用

在RS模型中,我们的研究结论指出[SES]与[数学成绩]的关系在各校之间并没无显著不同。因为[SES]斜率的随机性未达到统计显著水平,导致我们无法通过学校的特质,例如[公立]与[生师比]来解释[SES]在学校间的差异,因为学校差异并不存在。相对的,在两个CWC模型中,[SES]的斜率方差具有统计显著性,因而可以进一步探讨[SES]与[数学成绩]的关系在各校间的差异是否可以被学校特质变量来解释。因此,我们可以提出一个新模型,把[公立]与[SES]的跨层级交互作用、[公立]与[家庭作业]的跨层级交互作用纳入CWC模型中。

表5.2 RS与CWC模型的跨层级交互作用

	RS		CWC(N)		CWC(M)	
	EST	SE	EST	SE	EST	SE
截距	54.00	0.57	60.30	0.68	48.74	0.65
[家庭作业]	0.86	0.10	N/A		N/A	
[$\underline{\text{家庭作业}}$]	N/A		0.76	0.10	0.80	0.10
[SES]	4.30	0.09	N/A		N/A	
[$\underline{\text{SES}}$]	N/A		2.96	0.26	2.97	0.26
[公立]	-3.35	0.37	-5.57	0.38	0.44	0.29
[生师比]	-0.16	0.02	-0.29	0.03	-0.06	0.02
[家庭作业平均数]	N/A		N/A		1.62	0.20
[社经地位平均数]	N/A		N/A		7.98	0.25
[公立]×[家庭作业]	0.48	0.11	1.65	0.11	0.51	0.11
[公立]×[SES]	N/A		1.03	0.28	1.03	0.28
偏差度 Deviance	153 272		153 853		152 963	

新模型以RS,CWC(N)与CWC(M)三个模型进行分析的固定效果结果列于表5.2。在这个模型中,不论是以原始分数或是平减分数形式,[SES]与[家庭作业]仍然为学生层次的解释变量。学校层次的解释变量则是虚拟的[公立],另外增加了一个反应各校班级规模的[生师比]变量。三个模型中,除了RS模型,对于[SES]与[家庭作业]两个变项都设定为具有随机截距与随机斜率。在RS模型中,

[SES]具有固定的斜率,因为在表5.1中,显示[SES]的斜率的随机性并没有统计的显著性。同样的理由,表5.2中的RS模型并没有与[SES]有关的跨层级的交互作用。

分析的结果再一次指出变项中心化的影响。当第二阶层没有平均数置回的模型,包括RS与CWC(N)模型,强化了[公立]与[生师比]两个学校层次解释变量的影响力。在RS模型中,[公立]的效果很强,估计值为-3.35,z值为9.05;在CWC(N)模型中,[公立]的效果甚至更强,估计值为-5.57,z值为14.7。但是在CWC(M)模型中,[公立]的效果却变成正向,但是没有达到统计的显著水平,z值只有1.52。此外,另一个学校层次的解释变量[生师比]的系数强度,也有类似的情况,在RS模型中,强度为-0.16,z值为8,在CWC(N)模型中则提高到-0.29,z值为9.67,但是到了CWC(M)模型又再次降低,强度为-0.06,z值为3。这些学校层次解释变量系数数值的变化,说明了从CWC模型中增减平均数会影响学校层级的系数。

数据的操作如平减与模型中平均数的增减都会产生影响。对不同模型的比较可以看出,这些操作对于学校层次效果的影响会引出不同的结论,但是在学生层次,影响则有限,在各模型间得到类似的结论。

我们还有一个问题并未回答,亦即哪一个模型才是正确的,这个问题并没有办法单从技术层次上来回答,因为表5.2的三个模型都是"正确的"。如何选择,必须就研究者本身对于数据本身的理解与理论知识,以及研究的目的来考虑。如果研究者感兴趣的是模型能够解释反应变量的变异量的多寡,而不是在于第二阶层的效果,那么利用RS模型来拟合观察数据是最简单的做法。研究者不必去担心如何处理平均数的问题,因为一开始平均数就没有被中心化处理。同时,如果研究者较关心的是个别学生的表现,而不是学校间的差异时,RS模型也是最好的选择。

CWC模型的选择,则在于研究者想要把学生层次与学校层次的效果加以分离这层考虑,因此,有两个模型需要分别检测,即个别的学生模型与整体的学校模型。在这种情况下,中心化的决定以及是否置回平均数,就必须基于理论来考虑。以我们的例子来看,如果理论指出[SES]与[家庭作业]也是重要的学校层次的特征,那么我们就必须考虑是否要在比较不同类型的学校(公私立)之前就先把各校的差异控制住,当然,不同的做法所得到的结果就会不同。

RS,CWC(N),CWC(M)三个模型的选择由研究者决定,我们的例子说明了此一选择具有关键性的影响,而且,这个决定很难抉择。从技术的角度来看,中心化是不错的选择,因为它可以去除随机截距与斜率的高相关,以及第一层、第二层与跨层级交互作用之间的高相关(参见第5.7节)。中心化使得模型较为稳定,使得研究者所报告的系数多少具有独立性。从统计学的观点来看,我们知道平减模型

通过去除第一(学生)与第二(学校)层级变量的相关,成为较稳定的模型。但是在另一方面,中心化或许能够让一个并不符合研究需要的模型变得更拟合,但这似乎一点意义也没有。

正如我们在第2章所讨论的,脉络模型可以在平减或未平减的情况下达到良好拟合。那时的讨论是在固定系数模型的情况下探讨总方差如何被切割成不同的部分。但是对于脉络效果的解释,会因为我们有没有对数据进行操作而有所不同。在第二章中,原始分数模型的脉络效果是以 $b_W - b_B$ 来定义,而平减后的脉络效果则为 b_B。在 RC 模型中,平减会得到一样的效果,另外加上随机系数的效果。在固定及 RC 模型中,对于是否进行中心化平减程序的取决原理也是一样:在不同的情况下有所不同,如果一旦决定要进行平减,那么本书的建议是把平减用的平均数置回模型之中。如果没有这么做,研究者用 CWC(N)模型来进行估计,那么所估计得到的效果并没有控制住学校间差异,这个效果没有控制平减后的第一层次解释变量的平均数效果,然而在成长曲线模型中,这种效果恰好是研究者所要的,因为成长曲线模型的解释变量是时间。

通过这些有关中心化影响的讨论,以及通过实际数据与真实研究问题的演示,我们希望引起研究者对于这个问题的重视,我们的数据提供了一个理由,促使研究者去认真思考中心化不仅是在提供统计上的稳定性而已。如果平减程序被正确使用,可以让研究者获得更明确的、不被混淆的参数估计结果的回报,这个回报在固定系数模型或随机系数模型中,都是一样会很明显的!

5.3 模式变异

常见问题二:解释变异量或 R^2 在阶层线性模型中是如何定义的?阶层线性模型是不是有与 R^2 类似的概念?我们可不可以说结果变量的百分之多少的变异归因于群体因子,多少百分比的变异归因于个体因子?如果我们增加或减少变量,所估计的统计数据会有如何的改变?

从使用者的观点来重新定义多层次模型当中的多元相关平方(R^2)与解释变异量(explained variance)(注3)是很有意思的。但是很不幸的,这并不是一个可以简单回答的问题,因为有关组间方差有两种不同的定义:τ^2 与 ω^2,前者在本书中已有相当程度的讨论,后者则是由 Snijders 与 Bosker(1994)所定义。以两种的其中一种对于组间或组内方差的估计都会得到不同的结果。我们并不想花费篇幅来讨论这两种方法以何者来估计 R^2 会较佳,因为 R^2 的应用在多层次模型受到诸多限制,只有在经过中心化后且带有随机截距的模型中,R^2 的概念才能被清楚地定义。

之所以会有限制的主要理由，是因为在具有随机斜率的模型中，方差无法进行独立完整的估计。我们的讨论是建立在 Snijders 与 Bosker(1994)的观点之上，也就是模式变异与中心化有关。

在一般的回归分析(OLS)中，并没有对于组间差异进行估计。一个具有单一解释变量 x_{ij} 的标准回归模型如下：

$$\underline{y}_{ij} = a + bx_{ij} + \varepsilon_{ij} \tag{5.6}$$

误差项 ε_{ij} 的平均数为 0，方差为 σ^2，参数 σ^2 的最大似然估计值就是残差方差。换言之，围绕在回归线周围的残差平方和就是残差变异。

我们应该很熟悉回归当中的方差分析摘要表，归因于回归的 SS(平方和)与归因于残差的 SS 相加后得到结果变项 y_{ij} 的总离均差平方和，在第 2 章所介绍的脉络模型也有相同的概念，因为虽然解释变量不太一样，但是估计方法都是用 OLS 法来处理。在这些例子中，模型中仅有一个与方差有关的成分项，反应了残差平方和。在这种情况下，当模型具有明确的模式变异或多元相关平方的概念时，才能计算出回归模型的解释变异量。

但是，尽管在前述最简单的想法下，如果我们要宣称有多少变异是被哪一个解释变量所解释，则会受到整个多层次模型的分析原理以及所使用的名词与术语的诸多限制而遭遇困难。当解释变量之间没有相关时，模式变异是回归系数的平方和(这种状况仅存在于各实验组具有相等受试者的实验研究当中的解释变量)，此时解释变量的回归系数的平方和，即是模型中导因于该解释变量的模式变异(modeled variance)。但是这一个说法无法套用在当解释变量之间具有相关的情形，不幸的是，解释变量之间通常很难没有相关，就好像心理学上对于"基因"能够解释多少，"环境"可以解释多少，或是种族与社经地位各可以解释多少的这种无止境的辩论。社会科学家对于这些争议其实提出了一些解决之道，但是都是在回归分析的架构下所提出，而且并不是十分具有说服力。以下，我们可以看到如果应用了多层次模型来进行参数估计，或是多层次模型是正确的模型，那么这些问题就会变得非常复杂。

5.3.1 随机截距模型

到了现在，大家对于随机系数模型应该已经非常熟悉，加底线的变量是随机变量，模型的公式如下：

$$\underline{y}_{ij} = \underline{\alpha}_j + \beta x_{ij} + \underline{\varepsilon}_{ij} \tag{5.7a}$$

其中截距与第二层的解释变量 z 有关：

$$\underline{\alpha}_j = \alpha + \gamma z_j + \underline{\delta}_j \tag{5.7b}$$

这两个方程式中都带有一个误差项,分别是第一层解释变量的扰动项 ε_{ij}(方差为 σ^2)与第二层解释变量的扰动项 δ_j(方差为 τ^2)。如果这两个方差可以加总得到有意义的总变异,我们的问题或许得以解决。但是很不幸的,这两个方差分别处于两个层次,会有相互混淆的问题,我们可以在下面有关变异性的讨论中看到这里面的问题。

在一个模型中,偏差度(deviance,以 Δ 表示)的公式如下:

$$\Delta = m\log\omega^2 + m(n-1)\log\sigma^2 + \frac{SSQ_B(\gamma,\beta)}{\omega^2} + \frac{SSQ_W(\beta)}{\tau^2} \tag{5.8}$$

上述公式中,m 是组数,n 为每一组内的观察值数(组内样本数),为了便于说明,我们假设各组为人数相等的平衡组设计。SSQ_B 是组间平方和,SSQ_W 是组内平方和,ω^2 的数值是组间变异的总合,称为总组间变异(total between variance),$\omega^2 = \sigma^2 + n\tau^2$。

为了简化,我们也假设所有的变量都经过了总平减的处理,所以我们可以拿掉公式 5.7 当中的 α(α_j 项的固定部分)。另外,我们以 $\tilde{x}_{ij}$ 与 $\tilde{y}_{ij}$ 表示经过组平减的变量,x_j 与 y_j 表示组平均数,我们可以试图解释为何方差会有混淆的现象,组内的平方和定义如下:

$$SSQ_W(\beta) = \sum_{j=1}^{m}\sum_{i=1}^{n}(\tilde{y}_{ij} - \beta\tilde{x}_{ij})^2 \tag{5.9a}$$

组间的平方和定义如下:

$$SSQ_B(\gamma,\beta) = n\sum_{j=1}^{m}(y_j - \gamma z_j - \beta x_j)^2 \tag{5.9b}$$

这个方程式所告诉我们的是第二层的回归系数 γ 仅作用在组间平方和,但是第一层的回归系数 β 则在两层都发生作用。这就是一种混淆(confounding),我们无法很顺利地将参数分离成组间与组内两部分。

公式 5.8 也显示了离异性的计算被变异成分加权,但是我们看到了更进一步的混淆现象,也就是公式 5.8 的第三项,组间平方和 $SSQ_B(\gamma,\beta)$ 具有 ω^2 的信息,ω^2 与 σ^2 及 τ^2 都有关,组间与组内平方和都有涉及 β 的信息(公式 5.8 的后面两项),在第 2 章我们已经讨论过,此一现象说明了回归系数 β 同时具有组内与组间的成分。因此公式 5.8 当中的 SSQ_W 与 SSQ_B 被回归系数 β 所混淆。如果第一层解释变量 x 经过组平减,这个混淆问题就会消失,因为所有的 $\bar{x}_j$ 是 0,导致 $SSQ_B(\gamma,\beta)$ 中没有 β 系数存在,β 系数的最大似然估计值 $\hat{\beta}$ 即为 b_W。另一种说法,我们在公式 5.2 中可以增加组平均数为第二层解释变量,亦即 $\bar{x}_j$ 会有自己的回归系数,因

此我们仍然可以让 $\hat{\beta} = b_W$。

在 β 与 γ 两个参数的最大似然估计值的推导之后，可以计算出变异成分的最大似然估计值，我们需要针对组间与组内变异进行定义：

$$\hat{\sigma}^2 = \frac{1}{m(n-1)} SSQ_W(\hat{\beta}) \tag{5.10a}$$

$$\hat{\omega}^2 = \frac{1}{m} SSQ_B(\hat{\gamma}, \hat{\beta}) \tag{5.10b}$$

前面我们已经定义了总组间变异为 ω^2，而 $\omega^2 = \sigma^2 + n\tau^2$，意味着截距变异（或组间变异）$\tau^2$ 的估计方法为：

$$\hat{\tau}^2 = \frac{(\hat{\omega}^2 - \hat{\sigma}^2)}{n} \tag{5.10c}$$

此一公式必须是 $\hat{\tau}^2$ 为非负数时才成立，但是如果发生了 $\hat{\sigma}^2 > \hat{\omega}^2$ 的现象，导致 $\hat{\tau}^2 < 0$ 的不合理情况，那会是一大麻烦。

5.3.2 使用虚无模型来计算 R^2

在第 4 章的分析当中，我们有时候会在虚无模型中定义所谓的多层次模型的 R^2 或被解释变异量。虚无模型（null model）是只有随机截距的模型，截距的误差项涉及了第二层误差项（$\underline{\delta}_j$）与第一层误差项（$\underline{\varepsilon}_{ij}$）两个层次。因此：

$$\underline{y}_{ij} = \alpha + \underline{\delta}_j + \underline{\varepsilon}_{ij} \tag{5.11}$$

误差项的方差分别为变异成分 $\hat{\tau}_0^2$（反应的是组间方差）与 $\hat{\sigma}_0^2$（反应的是组内方差）。在平衡组设计的情况下，$\hat{\tau}_0^2$ 反应的是 SSQ_W，而 $\hat{\sigma}_0^2$ 反应的就是 SSQ_B。

如果更进一步将随机截距模型增加变量，我们可以发现新的变异成分估计值，$\hat{\tau}_1^2$ 与 $\hat{\sigma}_1^2$，如果把前面虚无模型的变异成分估计值减去新的变异成分，可以发现误差项降低了。换言之，我们预期通过增加微观层次的解释变量，可以减少学校内的变异，然而宏观层次的解释变量仅能降低学校间的变异。在某一个层次的误差变异的减少可以利用削减百分比的形式来表现，也就是解释变异量的数值（R^2）。在虚无模型中，两个 R^2 的计算，可以利用公式 5.12 来求出：

$$\hat{R}_B^2 = \frac{\hat{\tau}_0^2 - \hat{\tau}_1^2}{\hat{\tau}_0^2} \tag{5.12a}$$

$$\hat{R}_W^2 = \frac{\hat{\sigma}_0^2 - \hat{\sigma}_1^2}{\hat{\sigma}_0^2} \tag{5.12b}$$

公式显示了两个 R^2 的计算都是拿虚无模型的方差减去新模型的方差，也就是将与原来的虚无模型相比较的差异作为削减的比例。然而这个简单的公式有其限制，因为有可能得到负的 R^2，也就是公式 5.10c 当中 $\hat{\tau}^2 < 0$ 的不合理情况。在第 1 章我们也曾经提到，在随机斜率模型，我们并不采用这个公式。

5.3.3　使用总组间变异

基于 Snijders 与 Bosker(1994)对于总组间变异(total between variance; $\hat{\omega}^2$)的定义，可以阐明增加解释变量对于组内与组间变异的影响。首先，我们增加一个第二层解释变量，此时并没有组内变异的变动，因此 $\hat{\sigma}_1^2$ 与 $\hat{R}_W^2$ 维持不变，但是总组间变异 $\hat{\omega}^2$ 会减低，而 $\hat{\omega}^2 = \hat{\sigma}^2 + n\hat{\tau}^2$，因此 $\hat{\tau}_1^2$ 也会减少，亦即 $\hat{R}_B^2$ 会增加，对研究者来说是很正面的结果。

如果我们增加一个不影响组间变异的第一层解释变量(例如经过组平减的变量)，$\hat{\sigma}_1^2$ 减少而 $\hat{R}_W^2$ 增加，这也是正面的结果。但是 $\hat{\omega}_1^2$ 保持不变，意味着 $\hat{\tau}_1^2$ 会增加，公式 5.12a 所定义的 $\hat{R}_B^2$ 会减少，这就很奇怪了。我们增加一个变量，不影响组间变异，但是组间解释变异量却降低了。这又是因为混淆效应的影响。

如果我们改以下式来定义 $\hat{R}_B^2$：

$$\hat{R}_B^2 = \frac{\hat{\omega}_0^2 - \hat{\omega}_1^2}{\hat{\omega}_0^2} \tag{5.13}$$

若使用总组间变异，$\hat{R}_B^2$ 就不会改变，这样还比较有意义。这就是 Snijders 与 Bosker(1994)所主张的概念。

一般来说，变项有组间与组内变异两个部分，因此增加变量会使 $\hat{\sigma}_1^2$ 与 $\hat{\omega}_1^2$ 都变小，但增加一个对于组间变异影响很小的变量，多会提高 $\hat{\tau}_1^2$，进而造成组内相关的提高。

5.3.4　结　语

这一节当中，我们说明了增加了解释变量后，多层次模型误差变异可能会增加，这个现象违反了我们的直觉，因为传统上的理解是增加变量会降低(或不影响)误差变异，会有 SSQ_B 或 SSQ_W 降低的这个现象，在本节已经利用公式来说明。公式的推导说明了当 $\hat{\omega}^2$ 维持不变的情况下，改变 σ^2 与 τ^2 的一部分会影响 σ^2 与 τ^2 的另一部分。SSQ_W 与 SSQ_B 的混淆效果可以通过解释变量的组平减，然后将平减的平均数作为第二层解释变量这个做法来避免。在 5.2 节中已经详细地说明了这一种数据处理方法对于结果解释的影响。一般而言，我们会建议在计算 $\hat{R}_B^2$ 与 $\hat{R}_W^2$ 时不要

让公式内容太过于复杂,这两个参数在定义上具有争议且非常模糊,尤其在随机斜率模型的应用上有其限制。

5.4 统计检定力

常见问题三:如果我们使用多层次线性模型,回归分析的检定力如何说明?

这个问题无法以一般通则性的答案来响应。本节所列举的检定力函数只是多种可能中的几种。抛开其他的层面不谈,函数的形式受到虚无模型的一些条件的影响。前面已经提到了,虚无模型将方差切割成组内与组间两个部分,可能会有高度或低度的组内相关。若以第 1 章的表 1.1(Barcikowski 的研究结果)来说明,将团体的规模与组内相关强度一起考虑,对于拒绝虚无假设的统计检定力会有不同的影响(α 水平假设为 0.05)。

统计检定力是正确拒绝虚无假设的机率,因此检定力与虚无假设有关,而且是当虚无假设为伪时的特殊情况。在此一情况下,有关虚无模型的一些假设都可能会违反:

- ε_{ij} 为独立
- ε_{ij} 为平均数为 0,方差为 σ^2 的常态分配
- $\underline{\delta}_j$ 亦为独立且平均数为 0,方差为 τ^2 的常态分配
- $\underline{\delta}_j$ 与 ε_{ij} 相互独立

同时,检定力曲线可能会因为估计方法的不同、ICC 强度的不同、解释变量效果强度的不同、观察值数目的不同而有不同。

让我们从一个简单的模型开始,最简单的单一解释变量多层次模型是带有随机截距的模型:

$$\underline{y}_{ij} = \underline{\alpha}_j + \beta x_{ij} + \varepsilon_{ij} \tag{5.14a}$$

$$\underline{\alpha}_j = \alpha + \underline{\delta}_j \tag{5.14b}$$

为了讨论一项特殊效果的检定:β 效果的检定统计检定力,我们需要对于理论模型与真实模型的拟合性和 β 系数的估计方法加以定义。因为统计检定力是用发现一个具有显著性的效果的机率来定义,此时该效果必须存在于模型中,且利用显著性检验来检验之。显著性检验非常倚重被估计的 β 系数的标准误的计算,因此 β 系数的估计抽样变异决定了回归模型的检定力。如果抽样误差估计出来的数值很小,检定力就会很大。如果抽样误差被估计出很大的数值,那么检定力就会很小。

我们可以利用三种 $\hat{\beta}$ 参数抽样方差估计值来证明。第一种是最大似然法的方差,以 $V(\hat{\beta}_{ML})$ 表示。第二种是最小平方估计,以 $V(\hat{\beta}_{OLS})$ 表示,此种方法会对于多层次模型的方差进行校正,因为它把组内相关纳入考虑。第三种估计值是不处理组内相关的传统回归分析(OLS),此时 OLS 的方差估计值为 V_{OLS} 。在此我们并不想列出这些估计方法的公式,因为大家不会有兴趣去花费这个心思。但是我们可以说明,"不正确的"OLS 的方差估计值为 V_{OLS} 有最小的数值,而"正确的"OLS 的方差估计值 $V(\hat{\beta}_{OLS})$ 数值最大,最大似然法的方差 $V(\hat{\beta}_{ML})$ 则居中。

$$\hat{V}_{OLS} \leqslant V(\hat{\beta}_{ML}) \leqslant V(\hat{\beta}_{OLS}) \tag{5.15}$$

在此可以得到一个结论,使用"不正确的"OLS 估计值 V_{OLS} 因为有最小的标准误,因此会得到最高的检定力,但是这并不是适当的估计法。如果我们认为组内相关不存在,我们"低估"回归系数的标准误,我们较可能得到显著的检定结果,同时我们可能会错误地拒绝虚无假设,在表 1.1 的 Barcikowski 研究数据已经证明了这一点。我们再以图 5.1 来说明。

在此我们区分出三种可用于显著性检验的标准误:

1. 不处理组内相关的传统回归分析(OLS)标准误。在检定时,我们若用 $\hat{V}_{OLS}$ 来求出标准误是错误的(如果带有第一层解释变量的多层次模型为真),除非假定组内相关为 0。

2. β 系数 OLS 估计值的标准误 $V(\hat{\beta}_{OLS})$ 。此一标准误是正确的标准误,因为它把组内相关纳入考虑。

3. β 系数的最大似然估计值的标准误 $V(\hat{\beta}_{ML})$,也是正确的标准误形式。

5.4.1 范 例

在本节当中,我们将讨论一些简单的显著性检验的检定力问题。我们以使用虚无假设 $\beta = 0$ 的单尾检验为范例,因此在 α 水平为 0.05 时,如果所计算出来的 $\hat{\beta}$ 除以标准误大于 α 水平下的临界值 $z(\alpha)$,根据一般的常态理论下的 z 值为大于或等于 1.96,我们即可拒绝 H_0。在真实的情况下,我们可能以错误的标准误来进行检验。更清楚地说,如果我们以最大似然法来计算 $\hat{\beta}$,正确的标准误则是 $V(\hat{\beta}_{ML})$ 的平方根,如果我们以 OLS 法,我们应该求 $V(\hat{\beta}_{OLS})$ 的平方根。但是我们也可能以 OLS 法求出 $\hat{V}_{OLS}$ 的平方根为标准误,依据公式 5.14,在阶层模型中,这就不是正确的标准误。我们拒绝 H_0 的机率(检定力),亦即 $\hat{\beta}$,除以标准误,会大于 $z(\alpha)$ 。

以三种不同的标准误来进行显著性检验所得到的拒绝虚无假设的机率,可得到三条函数,列于图 5.1。这些函数是以单尾检定、组内相关为 0.5($V_B(x)$ 与

$V_T(x)$ 的比值)，$\omega^2 = 10$，$\sigma^2 = 10$，$\alpha = 0.05$ 等特定条件所得到的检定力函数。水平轴为 β 的真值，垂直轴为在 β 真值下拒绝虚无假设 $\beta = 0$ 的机率。

图 5.1 的解读如下：虚无假设的拒绝，发生在当真值大于 0（单尾检定）的情况下。当虚无假设必须被拒绝时（即真值大于 0 时），拒绝虚无假设的机率越大，该方法的检定力越大。另一个同样重要的问题是三种方法的偏误，当 β 的真值为 0 时，错误拒绝虚无假设的机率越高，偏误越大。这三条曲线反应了前面我们所讨论的三种状况。实线是使用错误的 OLS 法估计值的检定力，短虚线是使用正确的 OLS 法估计值的检定力，点虚线是使用最大似然估计值亦具有正确标准误的检定力。

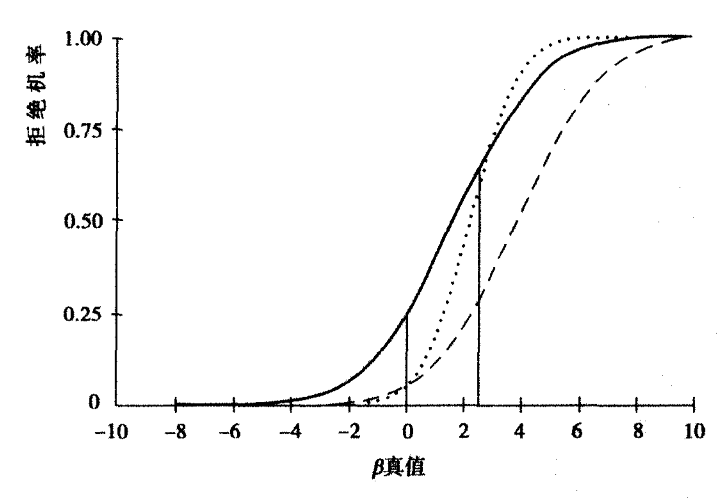


图 5.1　检定力函数图（拒绝机率为 β 系数的函数）

图形显示当 β 的真值越大时，这三种方法拒绝虚无假设的机率都会增加，这三种方法与 β 真值的关系具有一致性，但是函数的变动性是不一样的。图中以 OLS 法得到的错误标准误的检定力函数的那条实线，在 β 真值为 0 时，拒绝虚无假设的机率过高。在这里及在表 1.1 中都可以看到，使用错误的标准误拒绝虚无假设的机率大约为 0.25，而非正确的 0.05，因此可以说这个检定是非常偏颇的检验。我们太急于拒绝而没有保护到虚无假设。

同时，使用错误标准误的那条实线的检定力，在 β 的真值为 0 到 2.5 时（β 为真但是不高），拒绝虚无假设的机率都是最高的。当 β 的真值到达 2.5 时，以最大似然法求得的检定力点的虚线，拒绝虚无假设的机率最高。而短虚线（以 OSL 法估计的正确标准误）虽然是不偏的（当 $\beta = 0$ 时，$\alpha = 0.05$），但在三种状况中具有最低的检定力。

整体来说，最大似然法检定力曲线具有最佳的结果。首先，它具有正确反应第一类型错误的能力，第二，它能最快达到机率为 1.00 来拒绝虚无假设的检定力。

使用正确的标准误的 OLS 估计值则有最低的检定力，在整个真值范围内所拒绝的虚无假设都是最少的一种估计法。最大似然估计法即没有偏误，也有接近完美的检定力。

当各组的规模很小且组内相关很高时，使用错误的标准误的 OLS 估计法会更容易错误拒绝虚无假设。图 5.2 更清楚地说明此一现象。图 5.2 是当 $\beta=0$ 的虚无假设为真，用 V_{OLS} 计算出的错误标准误进行检验的结果。我们绘出了错误拒绝虚无假设的机率在两种不同的组样本人数与不同组内相关强度下的变化。在图 5.2 中，组内相关的大小放在横轴，错误拒绝虚无假设的机率放在纵轴，在单尾检验的情况下，最高的拒绝机率为 0.50。实线为组样本人数为 10 的情况，短虚线为组样本人数为 100 的情况。图 5.2 的两条曲线与第 1 章的表 1.1 的趋势相同。当组内相关大于 0.5 时，α 水平会到达最大值（接近 0.50），尤其是当组内样本数很大时。实线表示组内样本数较小者（本范例为 10），错误拒绝虚无假设的机率会比组内样本数大者低。图 5.2 显示，使用不正确的标准误所进行的显著性检验，会过度容易拒绝真实的虚无假设，即使在中度的组内相关的情况下也如此。在双尾检定的情况，或以 t 检定检验变异成分时，结果也大致相同。

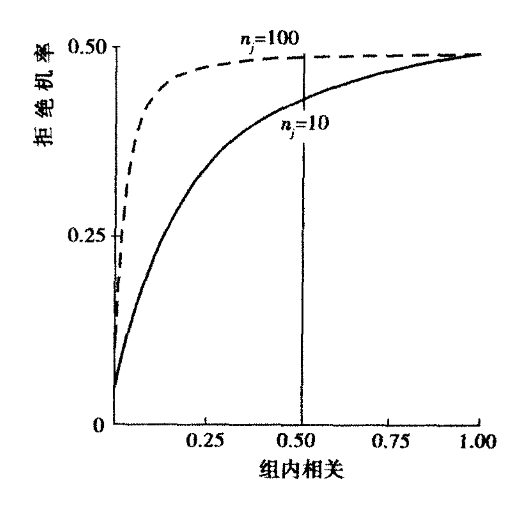


图 5.2 第一类型错误与组内相关的函数关系

5.4.2 模拟研究的发现

蒙地卡罗法（Monte Carlo）是另一种用来探讨多层次模型的检定力的方法。许多不同的参数都可以拿来检验，每一种状况都反应了一种“理想上”的最大检定力。我们可以区分出研究者所关心的两类参数：固定效果（γ）与随机效果或变异成分（τ）。摘述如下。

- γ 估计值：个体层次的斜率参数（γ_{10}）、截距参数（γ_{00}）、微观与宏观间的

跨层级交互作用参数(γ_{11}),宏观层次的截距参数(γ_{01})与斜率参数(γ_{11})。
- 变异成分:微观方差(σ^2)、宏观层次的截距方差(τ_{00})与斜率方差(τ_{11}),以及截距与斜率的协方差(τ_{01})。

有关 γ 估计值的研究,据我们所知有三个模拟研究(Bassiri, 1998; Kim, 1990; Mok, 1995)和两篇有关检定力的理论性论文(Snijders & Bosker, 1994; Cohen, 1995)。Bassiri 与 Kim 的两个研究表明在检验不同的组内相关、样本数、不同的组数下的 γ 估计值的检定力。另外有两个模拟研究报告了变异成分在不同的组内相关、不同的组数、组内样本数下的变化(Busing, 1993; van der Leeden, & Busing, 1994)。

Bassiri 的研究所使用的条件为:

- 两种不同的组内相关 0.10 与 0.25
- 10 到 150 的不同组数
- 5 到 150 的不同组内样本数

Kim 的研究则以第二次国际数学研究(the Second International Mathmatics Study; SIMS)的数据进行再次检验,他的模拟研究随机选择 50 个样本,检验不同条件的影响:

- 不同的关系型态:低强度、高强度、混合强度模型
- 25,50,100,200 四种不同组数
- 10,20,40 三种不同组内样本数

模拟研究所得到的结果非常接近。两个研究的结论指出,第一层估计值的变化原理与第二层估计值有所不同,也与跨层级交互作用的估计不同。第一层估计值为了达到理想的检定力,总样本数是重要的因素。此一原则也适用于传统的线性模型,但是当组内相关很高时,多层次模型需要较大的样本数来保证。第二层估计值的检定力大小则取决于组数的多寡。当问及我们到底是需要较多的组数,还是各组当中需要较多的观察值,另外有三个研究谈及这个议题。Snijders 与 Bosker (1994)计算了增加组数与检定力的消长关系,以及所耗费的成分问题。研究结果指出,增加组数比增加组内样本数要耗费许多倍的成本。另外两个研究也探讨了类似的问题(Cohen, 1995; Mok, 1995),但是可惜的是都没有正式发表。

Cohen(1995)的研究检验了随机增加一个学校或在一个学校中随机增加一个学生的影响,然后以近似函数来估计变异成分的标准误或回归系数的标准误,以求得每个学校理想的抽样人数。在多数的情况下,此一数目是组间与组内变异成分的比值的平方根。此一比值决定了研究的成本。

在 Mok(1995)的研究中,模拟的发现如下:

> 与过去典型的丛集抽样设计的研究文献一样,本研究发现如果研究者的能力能够获得 n 个样本,以在 J 个学校各取样 I 个学生来表示,那么以较多的学校数(较大的 J)与较少的学生(较小的 I)会比较少的学校数(较小的 J)与较多的学生(较大的 I)的偏误较小,效能更好。

Bassiri(1988)与 de Leeden & Busing(1994)分别对于跨层级交互作用的检定力进行探讨。这两个研究均指出,为了获得理想的跨层级交互作用的检定力,至少要有 30 个组,每组要有 30 个观察值。同时,研究也发现当有 60 个组,每一组有 25 个观察值时(总样本为 1 500),会得到相当理想的检定力。在较少组数的时候,例如 30 组时,每一组需要更多的样本以获得 0.90 的检定力。当组数较大的时候,例如 150 组时,每一组只要 5 个观察值就可获得 0.90 的检定力,此时的总样本数为 750。当组数或组内样本数较少时,对于跨层级交互作用的检定力会快速降低。为了获得 0.90 以上的跨层级交互作用检定力,Bassiri 发现多选择一些组,会比在各组内多选择一些观察值有帮助。

这两个模拟研究对于变异成分在不同条件下的状况时,发现与估计所使用的算法方法无关(IGLS 或 RIGLS),其变异成分会有低估与偏误的现象,偏误在 300 个组以上时不会发生。当抽样的组数少时,例如 5 或 10,得到的参数估计会有很大的变化,且没有包括真值。有关这些研究,请参考第 5.6 节的讨论。

5.4.3 结 语

有关检定力的差异,系发生在固定效果的变异成分。经本节的说明,我们的结论是最大概似估计法具有最高统计检定力。但是因为固定效果不受估计方法的影响,而标准误会受影响,我们的结论是在多层次模型中,对于固定效果的估计,以 EB/ML 估计法平均而言会有最高的检定力和最小的第一类型错误犯错率(参见图 5.1与图 5.2)。

一般来说,为了获得研究者所需的足够检定力,观察值要多,除非研究者所探究的现象有非常强且容易被侦测到的效果。然而,研究最理想的样本数为何,每一个研究都有所不同。当组数很少时,随机成分会被低估(在 IGLS 时),或有较大的标准误(在 RIGLS 时)。对于跨层级效果,要有足够检定力,组内样本数不能太少,且组数要大于 20。

有关检定力的问题,非常容易受到效果强度的影响。对于各参数,乃至于组内相关皆然,尤其是第二层的估计值与跨层级的交互作用。在本节一开始就已经提到,第三个经常被问到的问题很难一言以蔽之。当然还有更多的状况可以分析,但是当我们检验越多的条件,得到的结果也越趋复杂而模糊。模拟研究的结果指出,

固定效果的结论远比随机效果来得直接、清楚。基于其他的理由，例如研究的成本效益与取样可能性的考虑，有关较多的组数与较多的组内样本数哪一个重要，是一个较有意义的问题。前面所举出的三篇论文都提到了这个问题，读者可以参考之。

5.5 随机问题

常见问题四：多层次分析当中的系数可以是非随机且对于各组都是固定值，也可以是非随机但是在各组间是变动的，也可以是随机的（也就是自成一个变量）。这些不同的作法应如何选择？不同的选择之间有何差异？

也是一样的，这个问题的答案，也是随着不同的状况而有不同。如果模型中带有随机斜率，或是模型中仅有一个随机截距时，状况都不尽相同。在模型中，组间与组内回归系数的相对大小，微观层次方差与宏观层次方差的比例，也都会影响这个问题的答案。为了简化说明，我们将说明增加一个随机截距或随机斜率后，对于固定效果估计的影响。此一改变的影响可大可小，取决于第一层方差 σ^2 的大小与第二层方差 τ^2 的大小。

5.5.1 ANCOVA、RANCOVA 与简单回归

首先，我们来看看协方差分析，模型为：

$$\underline{y}_{ij} = \alpha_j + \beta x_{ij} + \underline{\varepsilon}_{ij} \tag{5.16}$$

上式看起来很类似于一般回归模型（OLS），从截距的下标 j 来看，我们可以预期各组的截距有所不同。而由于截距为平均数（在此处应为调整后平均数），各组平均数是假设有所不同。在一个 ANCOVA 模型，平均数的比较是主要的焦点，因此每一组有其自己的截距，而截距是被设定为固定的，数值未知。所有的组都有相同斜率 β 的回归线。

相对应的随机系数模型，称为 RANCOVA，看起来非常相似，所不同的是截距并不只是变动的数值（如公式 5.16），而且是随机的。在公式 5.17 中，α 项带有底线。截距是变动的或随机的，不仅具有统计上的重要性，也有理论上的意义。RANCOVA 的模型如下：

$$\underline{y}_{ij} = \underline{\alpha}_j + \beta x_{ij} + \underline{\varepsilon}_{ij} \tag{5.17a}$$

$$\underline{\alpha}_j = \alpha + \underline{\delta}_j \tag{5.17b}$$

ANCOVA 与 RANCOVA 最重要的差异在于概念层次，公式 5.16 的 α_j 项可为任何一组数值，他们可能非常的不一样，也可能是二分的数值，或是其中的截距特别

大,其他的差不多,诸如此类。相对的,公式5.17a的α_j项,则假设是来自一个平均数为α,方差为τ^2的常态分配母体中的一个随机样本。此一假设使得截距项数值具有某种规律,尤其是当组数很大时。

同时,我们也可以观察到一个现象,当宏观层次的方差τ^2为0时,公式5.17b当中即没有扰动项,而RANCOVA模型成为$\underline{y}_{ij} = \alpha + \beta x_{ij} + \varepsilon_{ij}$,如此即成为一个传统的线性回归模型而不是协方差模型,因为α为一常数,在组间不会有任何变化。

公式5.16与公式5.17相比较的结果,使我们可以发现ANCOVA的平均数结构非常清楚,每一组有一条回归线,而各回归线呈现平行(参见第3章的图3.1)。在RANCOVA,则仅有一条回归线,但是方差的结构较为清楚(参见第3章的图3.4),特别是在同一组的观察值相关时,结构更为明确。两者的另一个主要差异是在ANCOVA中,我们主要是以平均数来进行模型的分析,但是在RANCOVA中则主要是依赖方差与协方差的分析。这些差异也在第3章的第3.3节与3.4节的图形中可以看出来(注4)。

另外,比较ANCOVA与RANCOVA对于β系数的最大概似估计值的差异也是非常有意思的。Longford(1993)曾经进行详细的说明。我们在此仅作大略的说明。

为了解释的完整性,我们仍再一次与一般的OLS回归分析(仅有一个方差成分σ^2,各组的截距α_j均等于α)进行比较。如果我们从OLS回归转向ANCOVA,我们可以看到让截距可以自由变动后对于斜率的影响,如果再转向RANCOVA,我们可以看到让截距成为一个随机变量的影响。我们所说的所有的β系数估计值,都是组间回归系数b_B的加权平均数和组内回归系数b_W,在ANCOVA与RANCOVA模型中都是无偏估计值。

我们的起点在于典型的脉络方程式(公式2.7),我们再一次地列于下:

$$b_T = \eta^2(x) b_B + (1 + \eta^2(x)) b_W \tag{5.18}$$

$\eta^2(x)$是组内相关的平方,也就是组间方差所占的比例。此一公式告诉我们如果我们进行简单回归(也就当斜率估计值为b_T时)转向ANCOVA(也就当斜率估计值为b_W时),将会发生什么事。如果$\eta^2(x)$接近0,全体回归系数b_T与组内回归系数b_W将会接近相等。当组间平均数差异很小时,此一现象就会发生。同样的,如果b_B与b_W接近相等时,不论$\eta^2(x)$多大,b_T都会接近b_B与b_W。

当RANCOVA为平衡组设计时,每一组的观察值数目相等,最大概似估计的关系式会非常类似公式5.18。虽然因为涉及变异成分,关系式会显得较为复杂,但是结构基本上是一样的。两者都是b_B与b_W的加权平均数,权数为没有负值,且总和为1。斜率的最大概似估计值公式如下:

$$\hat{\beta}_{ML} = \frac{\lambda^2\eta^2(x)}{\lambda^2\eta^2(x) + (1 - \eta^2(x))}b_B + \frac{(1 - \eta^2(x))}{\lambda^2\eta^2(x) + (1 - \eta^2(x))}b_W \quad (5.19)$$

而

$$\lambda^2 = \frac{\sigma^2}{\sigma^2 + n\tau^2} \quad (5.20)$$

λ^2 是第一层方差 σ^2 与组间方差 ω^2 的比值。我们在第 5.3 节已经讨论过，$\omega^2 = \sigma^2 + n\tau^2$。

RANCOVA 的 β 系数最大概似估计值($\hat{\beta}_{ML}$)将介于 b_B 与 b_W 两个极端数值之间。如果所有的方差都存在第一层时，$\lambda^2 = 1$，也就是说，如果第二层的方差为 $\tau^2 = 0$，则 $\hat{\beta}_{ML} = b_T$。这是有意义的，因为我们刚刚提到，在这种情况下，RANCOVA 会变成带有一个固定截距 α 的简单线性模型。

如果组间方差与第一层方差 σ^2 相比之下大了许多，则 λ^2 接近 0，$\hat{\beta}_{ML}$ 会接近 ANCOVA 的估计值 b_W。如果组内样本数非常大时，就会发生这个状况，因为 $n\tau^2$ 会很大。

如果 $\eta^2(x) = 1$，也就是说，如果 x 是一个整体层次的变量而没有组内方差时，则 $\hat{\beta} = b_B$；相反的，如果 x 是一个平减后的个体层次变量而没有组间方差时，则 $\hat{\beta} = b_W$。

因此，对于如果把截距设定为随机会发生什么事这个问题的回答，仍然是很无奈的且不令人满意的一句话：视状况而定。让我们比较一下 ANCOVA 与 RANCOVA两者，如果 $n\tau^2$ 很大(也就是第二层方差或组内样本数很大，或两者很大时)，两者将不会有很大的差异，此时 λ^2 接近 0。公式 5.19a 中 b_B 的权数会接近 0，导致 RANCOVA 中的最大概似估计值会接近 b_W。同理，如果 $\eta^2(x)$ 很小，两种模型也不会有很大的差异。RANCOVA 若与简单回归有别，在 $\hat{b}_B$ 与 $\hat{b}_W$ 之间必定有相当程度的差异。但是即使 $\hat{b}_B$ 与 $\hat{b}_W$ 之间非常不同，若截距方差 τ^2 非常小，我们仍然会得到 $\hat{\beta}_{ML}$ 接近 b_T 的结果。

5.5.2 固定与随机斜率

如果随机斜率也考虑进来，事情就更不单纯了。此时本节的问题就变成了：让我们先估计斜率方差为 0(也就是斜率为固定值)的子模型中的斜率与截距，然后估计模型当中的斜率(但是不加上“让 β 为随机”这个条件)。此时截距与斜率的 $\hat{\alpha}$ 与 $\hat{\beta}$ 的估计会发生什么情况？

让我们先看一个带有一个解释变量的简单模型：

$$\underline{y}_{ij} = \alpha_j + \underline{\beta}_j x_{ij} + \varepsilon_{ij} \tag{5.20a}$$

斜率与截距都是随机：

$$\underline{\alpha}_j = \alpha + \underline{\delta}_{0j} \tag{5.20b}$$

$$\underline{\beta}_j = \beta + \underline{\delta}_{1j} \tag{5.20c}$$

为了简单起见，我们将一些条件加以限定使得模型成为平衡模型（注5）：我们假设每一组人数 n 相等，x_{ij} 为组平减后的数据型态，x_{ij} 变量的平方和（s_j）在各组也相同。进一步的，我们假设第二层方差 $\underline{\delta}_{0j}$ 与 $\underline{\delta}_{1j}$ 独立无关，方差分别为 τ_0^2 与 τ_1^2。虽然这些限制与重复量数的平衡设计不同，但这些限定使得模型显得很特别。我们仍然希望我们所推知的结果，即使在这些限定不存在的情况下，仍是真实有用的。

在我们限定这些假设的情况下，经过一些很厌烦的计算程序，我们可以发现 α 的最大概似估计值总是接近调整后平均数。当然，此一调整是以 β 的最大概似估计值来进行的调整，调整的方法近似于公式5.18与5.19，它提供了另一组 b_B 与 b_W 的加权平均数。

$$\hat{\beta}_{ML} = \frac{\lambda_0^2\eta^2(x)}{\lambda_0^2\eta^2(x) + \lambda_1^2(1-\eta^2(x))}b_B + \frac{\lambda_1^2(1-\eta^2(x))}{\lambda_0^2\eta^2(x) + \lambda_1^2(1-\eta^2(x))}b_W \tag{5.21a}$$

此时，我们得到两个组内与组间方差比值。一个是与随机截距有关的 λ_0^2，另一个是与随机斜率有关的 λ_1^2。

$$\lambda_0^2 = \frac{\sigma^2}{\sigma^2 + n\tau_0^2} \tag{5.21b}$$

$$\lambda_1^2 = \frac{\sigma^2}{\sigma^2 + n\tau_1^2} \tag{5.21c}$$

如果 $\tau_1^2 = 0$，也就是斜率不是随机，则公式5.21a的估计结果会与公式5.19a的结果相同。由这两个公式的比较，我们可以得知“让斜率为随机”会发生什么事情。

此时，我们可以进行与第5.5.1节相同的讨论。如果 $\lambda_0^2\eta^2(x)$ 比 $\lambda_1^2(1-\eta^2(x))$ 小很多，则 $\hat{\beta}_{ML}$ 会接近 b_W，读者若检阅公式5.21，可以发现何时这几个乘积项会很大，何时会很小。

如果 λ_0^2 与 λ_1^2 大约相等，则 $\hat{\beta}_{ML}$ 会接近 b_T。如果让斜率设为随机，λ_1^2 的估计值将会减小。此时如果其他的条件维持一定，表示 $\hat{\beta}_{ML}$ 会接近 b_B。

依照同样的逻辑,我们可以探讨如果"让斜率设为随机"时,回归系数的标准误会发生什么改变。但是与第5.4节一样,此时我们必须十分小心,因为标准误取决于两者:我们所假设的模型和真实模型。我们可能会提出一个并不正确的模型,例如将斜率设定为固定,但是事实上是随机。但是我们也可以假设一个模型带有随机斜率但是"事实"上是固定。因为固定是随机的一种特殊状况,也就是说它并不是"错误的"。但是我们可以预期,"让斜率设为随机"在两种情况下,会有非常不同的结果。

如果我们增加更多的变量,然后让其中一个变量的回归系数设定为随机,其他则否,读者会觉得整个问题变得十分复杂,这是一点都不令人讶异的反应。若要进行这类的讨论,我们就不得不利用矩阵代数,但这已经超过了本书所预期的目的。

5.6 估计方法与算法

常见问题五:什么是 FIML,REML,EM,IGLS,RIGLS,EB/ML,OLS,GLS?

Leeuw 与 Kreft(1995)的文章指出,对于模型(models)、技术(techniques)与算法(algorithms)三个概念进行区辨是很重要的一件事。

模型(model),特别是统计模型(statistical model),是由一组描述随机变量之间的关系的方程式所组成。值得注意的是,随机系数模型虽是在处理固定的预测变量,但是方程式中总是带着一个随机项(在本书中以划底线来表示),由扰动项与(或)随机系数所组成。模型中通常会有一组未知的参数(parameter),用以描述特殊的模型状态(instance)。如果我们的模型是个一元正态分布的随机变量,则模型的参数就是这个常态分配的平均数与方差。如同本章对于中心化(平减)的讨论所提及的,有时相同的模型(相同的一群随机变量),可用不同的参数化(parameterization)方式来描述其内容。

统计技术(statistical technique)是一种功能(或软件),可以把数据输入,制造出未知参数的数值。更广义来说,统计技术是将资料转换成一组统计数,这些统计数可以用来估计模型的参数,但是也可以用来描述统计量,或产生图或表。统计技术通常是通过将统计原理(statistical principle)应用到一个模型之上而发生功能。统计原理可能是最大概似法,或是最小平方法,诸如此类。如果模型是多层次这种特殊的模型,那么应用最大概似法的统计原理,就是要我们去计算模型参数的最大概似估计值。

技术是通过算法(algorithm)来实践。当我们决定使用最大概似估计值来估计模型参数,我们仍然可能使用不同的算法来进行运算。事实上,说得更明确的话,

我们可以说即使选择了某种算法，也不完全决定我们能做什么。在许多情况下，去比较使用相同算法的不同计算机软件是有意义的，就好像我们说相同的技术背后有不同的算法是一样的道理。

对于上述三个名词有所区分之后，我们可以开始来讨论一些在多层次分析常用的技术与算法。在第1章中，我们已经讨论了不同的模型与进行算法应用的不同计算机软件。

在多层次模型中，学者所取用的最重要统计原则是最大概似法(maximum likelihood)。然而，名词的使用有时会有一些混淆，因为人们可能取用相同的统计原理，但是处理方法略有不同。多层次模型把因变量以 $\underline{y}$ 来描述，然后应用最大概似法在模型参数的估计。$\underline{y}$ 的分布被假设为常态分配，平均数决定于回归系数(本书的符号为 γ_{st})，变异量数则决定于变异成分，我们以 ω_{st} 与 σ^2 来表示。这些参数有其相对应的技术来估计，简单来说就是最大概似法，但是有时也称为完全信息最大概似法(full information maximum likelihood; FIML)。

另外的替代作法，是我们可以将最大概似原理应用到最小平方残差的估计，称为限制性或残差最大概似法(restricted or residual maximum likelihood; REML)，意味着我们必须先移除固定变量的效果。记住，模型中的残差与固定效果是没有关连的。残差的分配也是常态，因为用 $\underline{y}$ 计算残差只是把加权的总和拿掉而已。但是残差的分配不再受到固定效果的估计值(γ_{st})的影响，仅受到变异成分的影响。因此，应用最大概似法来估计残差，意味着我们不能估计回归系数。这时，并不能尽符我们的需要，因此我们可以纳入其他的原则来估计回归系数。此时其他的原则为一般或加权最小平方法(generalized or weighted least squares; GLS)，其中我们使用估计变异成分去建立权数矩阵。

其他的原则也可以拿来应用，例如贝叶斯(Bayes)与特殊的经验贝叶斯方法(empirical Bayes)，在一个完全的贝叶斯取向中，未知的参数被视为是服从某一已知先验分布的随机变量，如此一来，贝叶斯回归就非常类似于随机系数回归。随机系数的分配遵循先验的分配。在完整的贝叶斯统计学中，我们运用先验的分配于贝叶斯定理中来计算参数的后验分配，其中导出十分复杂的计算过程，这个过程经常是通过马可夫链(Markov chain)蒙地卡罗算法(Monte Carlo algorithm)来完成，例如吉布斯抽样器(Gibbs sampler)。在经验贝叶斯方法中，我们并不假设一个先验的分配是完全已知，但是我们需假设它取决于一些已经被我们估计得出的未知参数。因此此一取向非常近似于随机系数或多层次取向。事实上，实证贝叶斯法基本上就等于是最大概似法。

如果今天算法是我们所关心的主题，那么就有很多选择了。HLM 软件(Bryk et al., 1996)使用的是 EM 算法来计算出 REML 的估计值和采用一些特殊步骤来

加速收敛(convergence)。EM 算法是一种普遍用来当数据中具有遗漏值(或随机参数)时来计算最大概似估计值的方法。这种特殊形式的概率提出一种逼近我们要求的最大化复杂函数的一个简单方法,在最大化这个简单函数的每个步骤中,我们会形成一个新的且较佳的逼近值。EM 是建立在这种必定会收敛;但是很费时的运算过程中。

Gauss-Newton 法在多层次分析中被称为计分法(method of scoring),在每次的迭代(iteration)过程中,需要另外进行一些工作,但是由于可以加速收敛的速度,因此需要较少的迭代。此一方法是建立在较接近概似函数的基础上,可以应用于 FIML 与 REML 两种方法。VARCL 软件(Longford, 1990)就是使用 FIML 法。两种方法都可以应用在 MLn 软件中,而 IGLS 是应用未受限的最大概似估计(FIML),RIGLS 应用受限的最大概似估计(REML)。后面两种软件是基于当如果我们知道变异成分后,概似函数只是简单地去估计最佳的回归系数,或是当如果我们知道回归系数后,概似函数去估计最佳的变异成分的这种计算过程中。因此,我们有两种代替这两种最小化的技术。第一是先猜测变异成分,然后以 GLS 法来求出最佳的回归系数。当给定了回归系数之后,我们可以以变异成分的二次函数来逼近概似函数。我们也可以用 GLS 进行最小化程序,来找出新的变异成分,诸如此类。

对于这些技术与算法的细节的了解超过本书的范围,我们希望读者不要被这些英文缩写给彻底打败,同时也能够区别 REML 与 RIGLS 等这些名词的不同。REML 是定义一个能够被最小化的损失函数(loss function),也就是一种统计技术;而 RIGLS 则是定义可求得特定损失函数最小化的一种方法。

5.6.1 FIML 与 REML 何者为佳

有关 FIML 与 REML 法的选择与某些原因有关。一方面是因为两种方法在 MLn 软件中都提供,其中 IGLS 是用来计算 FIML 估计值,RIGLS 是用来计算 REML 估计值,但是如何以及为何选择哪一种方法,则不是十分清楚。在 Bryk 与 Raudenbush(1992)的软件中,REML 有时表现得比 FIML 法好,尤其是较小的数据库。所谓小的数据库是指组数较少,而不是每一组人数少。REML 法是在固定效果被估计后才计算残差。Goldstein(1995)在他的书中讨论了两种方法的差异,但是也没有清楚地指出哪一种方法应该使用在哪一种情况下。只留下一些信息以及模拟研究的结果。从这些信息中,我们得知所有的估计方法中,对于变异成分估计最重要的条件是要有足够数量的组数。

因为蒙地卡罗的模拟是评估多层次估计方法的统计特性的技术,我们在此报告一些模拟研究的结果。

这些研究是基于对于某些参数进行人为操作产生的数据进行重复估计的结

果,因此并非完备的结果。多数的模拟是以低度组内相关($r \leq 0.25$)所进行,以符合多数社会科学研究的特性。两个研究探讨了固定参数估计值(γ)的特性(Kim, 1990; van der Leeden and Busing, 1994),两个研究检验变异成分在不同估计方法时的特性(Busing, 1993; van der Leeden, & Busing, 1994),后面两个研究的结果在第 5.6.3 节中描述,而有关 Kim 的研究设定则请参考第 5.4.2 节的介绍。

5.6.2 固定系数估计方法的影响

Kim 所比较的各种估计方法包括 OLS,GLS 与 EB/ML 法(REML 形式)。结果显示在各种条件之下估计的结果是相同的,同时研究结果也显示,若要获得不偏的估计值,不一定要用到像 EB/ML 法这种复杂度高的技术。他所比较的几种方法都可以获得不偏估计值。

我们重新计算 Kim 研究所获得的固定参数估计值的效能,来进行 OLS,GLS 与 EB/ML 三种估计法的比较。可以发现 GLS 与 EB/ML 法的 γ 估计值是相同的,在 OLS 与 EB/ML 法之间则有些微的差异,OLS 法的结果会有较大的估计方差。但是在大样本的情况下就不会这样,OLS 与其他两种方法在大样本时对于 γ 的估计并没有差别。OLS 法的效能在各种条件下都可以达到 0.90 的水平,若与其他方法相比之下有较低的效能的话,表示该方法要获得更多的观察值才能获得相同的效能水平。整个的结论大致如下:

- 在随机系数模型中,GLS 对于固定参数的估计可获得理想的结果。
- OLS 在估计随机系数的起始值的软件设定中,效能稍微低了一些。
- GLS 与 EB/ML 法具有相同的效能。
- 所有三种方法所得到的 γ 估计值都是不偏的。

Van der Leeden 与 Busing(1994)也进行了 OLS,GLS 与 RIGL(也就是 HLM 中的 REML 法)三种估计方法在处理跨层级估计值上的比较。研究结果就如同 Kim 的研究发现,这三种方法对于跨层级的 γ 估计都没有差别,而且这三种估计法都可获得不偏估计值。

5.6.3 变异成分估计法

Busing(1993)与 van der Leeden(1994)针对变异成分的估计方法进行模拟研究。前者是以非限定方法(IGLS or FIML)来估计变异成分,后者则是以限定方法(RIGLS or REML)来估计变异成分。两个研究都是使用 MLn 软件(Rasbash et al., 1990)来进行参数估计,估计方法是在不同样本大小、组内相关、组数等条件下来进行比较。研究的条件如下:

- 0.20,0.40,0.60,0.80 的组内相关

- 截距与斜率之间具有 0.25,0.50,0.75 的相关
- 样本数为下面有关组数与组内样本数等各条件的组合
- 组数:5,10,25,50,100,300
- 组内样本数:5,10,25,50,100
- 斜率、截距、斜率与截距的协方差等各变异成分以及标准误的相对不偏性的结果
- 每个状况都进行 1 000 次重复

变异成分以 IGLS 与 RIGLS 的估计比较发现,RIGLS 是比较不会产生偏误但效能较差的估计方法,因此很难说两种方法哪一种较为理想。同时估计的复杂度与估计效能之间的消长在不同的情况下也显得没有规则。这些研究的作者也无法对于何时该用何种方法给出一个明确的建议。

这两个研究指出,在第一次迭代后所报告出来的变异成分的 GLS 估计值,比起在收敛之后得到的估计值,精确度较低。对于 IGLS 与 RIGLS 两种方法,随着迭代的进行,各迭代的均方误降低,截距的变异成分估计值的精确度则会增加。相对于前面一节有关固定效果的研究,随机效果的估计会随着迭代的进行而会有所改善。

5.6.4 结　语

模拟研究的结果指出,以 OLS,GLS,REML 三种估计法来进行固定效果的估计都可得到不偏估计的结果。这三种方法在固定效果估计上的差异在于运作的效率。OLS 法在分析大型数据时,虽然可得到不偏估计值,但是效能较差。对于变异成分的估计,各估计法何者较佳并没有清楚的定论。一般来说,RIGLS(REML)会比 IGLS(FIML)偏误较小,但是 RIGLS(REML)的缺点是较不精确。

5.7 多元共线性

常见问题六:多元共线性在多层次分析中的问题有多严重?

在典型的回归分析中,多元共线性是一个非常重要的议题。在 Belsley(1991)的著作中有一完整的讨论。在多层次分析中,多元共线性也是非常重要的问题。在一个完整的多层次模型中,我们会有三种预测变量:第一层解释变量、第二层解释变量与跨层级交互作用(为第一层与第二层变量的乘积),这些变量可以区分为组间与组内两部分:

$$x_{ij} = x_{.j} + (x_{ij} - x_{.j}) \tag{5.22}$$

上式中的组平均数 $x_{.j}$ 与离均差 $x_{ij} - x_{.j}$ 彼此独立无关,第二层变量的组内变异

则为0。若第一层解释变量进行组平减后，组间成分亦为0，此时组平减后的第一层变量与第二层变量也是无关的。再者，如果第一层解释变量经过组平减，则与任何一个第二层解释变量的跨层级交互作用都无关。

如果我们在多层次模型中再增加一个跨层级交互作用，会造成何种影响值得观察。在第4章当中，我们已经看到交互作用会压制第一层的主要效果，导致一些不稳定的情况发生。

如果忽略多层次模型不只具有一个变异成分这个特性，我们可以套用增加变量来侦测跨层级交互作用增加的影响的一般理论（Weisberg，1985 第2.4节），亦即增变量理论（added-variable theory）来了解这个现象。

假设我们利用一组解释变量 x_1，x_2，…，x_p 来进行回归分析，若我们增加一个额外的变量 z，会有两个议题：

- 如果增加 z 变量后，β_1，…，β_p 会产生什么变化？
- z 变数的回归系数 γ 为何？

对于第一个问题的回答很简单，如果对于 y 以 z 变量作回归，计算出残差为 $\tilde{y}$。我们也对每一个 x 变量以 z 变量作回归，计算出残差为 $\tilde{x}_1$，…，$\tilde{x}_p$。然后我们可把 $\tilde{y}$ 以 $\tilde{x}_1$，…，$\tilde{x}_p$ 这一组变量作回归，计算出 $\tilde{\beta}_1$，…，$\tilde{\beta}_p$ 这一组系数。此时新的 $\tilde{\beta}$ 系数就是我们增加了 z 变量之后，从 $\tilde{x}_1$，…，$\tilde{x}_p$ 进行回归分析所得到的 β 系数。如果 z 变量是跨层级交互作用，意味着在进行回归分析之前，我们已经从第一层及第二层抽离了跨层级交互作用的效果。如果 z 变量与任一个 x 变量的相关很高，很明显，这个 x 变量的变异将会被大量移除，它的 β 系数值就会有很明显的改变。

第二个问题也可以利用相同的逻辑来分析，为了计算新增加的 z 变量的 γ 系数，我们可以把 y 以 $\tilde{x}_1$，…，$\tilde{x}_p$ 进行回归分析，计算出新的扰动项 y^+，然后把 z 变量也以 $\tilde{x}_1$，…，$\tilde{x}_p$ 进行回归分析，计算出扰动项 z^+，最后把 y^+ 以 z^+ 作回归，即可得到 γ 系数。如果 z 变量是跨层级交互作用，意味着在进行回归分析之前，我们已经从第一层及第二层抽离了与 z 有关的效果。如果我们假设在进行这项移除之后，已经没有什么有关 z 变量的信息被留下来，因此 z 变量的回归系数就会非常不稳定。

由于多层次分析具有一个以上的变异成分，因此前面所讨论的状况相对是很单纯的，第4章中即有一些复杂的模型，数学计算过程非常复杂，尤其是当我们把跨层级交互作用以随机项来处理之时，算式更是复杂。因此我们仅作上述的讨论，另外再以一些数据来示范。

在完整的 NELS88 数据库中，我们选择［家庭作业］与［SES］变量（在本节各以 H 与 S 来表示），然后以聚合程序计算出各校的［平均家庭作业］与［平均社经地位］两个第二层变数（各以 $\bar{H}$ 与 $\bar{S}$ 来表示）。我们可以得到四个跨层级交互作用项：

$\bar{H}H$，$\bar{H}S$，$\bar{H}H$，$\bar{S}S$。以[数学成绩]（以 M 表示）作为因变量，因此共有九个变量。这九个变项的相关系数矩阵列于表 5.3。

表 5.3 第一层、第二层与跨层级交互作用项的相关系数矩阵（下三角形内为原始分数相关矩阵，上三角形为组平减分数相关矩阵）

	H	S	$\bar{H}$	$\bar{S}$	$\bar{H}H$	$\bar{H}S$	$\bar{S}H$	$\bar{S}S$	M
H	—	0.11	0.00	0.00	0.97	0.12	0.05	-0.01	0.19
S	0.21	—	0.00	0.00	0.11	0.97	-0.01	-0.19	0.25
$\bar{H}$	0.38	0.34	—	0.52	0.00	0.00	0.00	0.00	0.30
$\bar{S}$	0.20	0.65	0.52	—	0.00	0.00	0.00	0.00	0.45
$\bar{H}H$	0.93	0.28	0.63	0.34	—	0.12	0.18	-0.00	0.18
$\bar{H}S$	0.23	0.96	0.39	0.66	0.33	—	-0.00	-0.06	0.24
$\bar{S}H$	0.21	0.52	0.48	0.81	0.40	0.59	—	0.07	0.01
$\bar{S}S$	0.07	0.02	0.21	0.18	0.16	0.19	0.29	—	-0.04
M	0.29	0.48	0.30	0.45	0.33	0.47	0.37	0.07	—

表 5.3 的下三角形中，[家庭作业]与[SES]变量为原始分数的数据型态，上三角形则是组平减的数据型态。很明显，在对角线以上的相关系数都比较小，很多相关系数都为 0 或接近 0。表中显示如果把第一层变量进行组平减处理后，多元共线性问题并不严重。但是我们仍可以看到非常高的相关系数，例如 H 与 $\bar{H}H$ 之间，S 与 $\bar{H}S$ 之间。这是因为[家庭作业]（H）的变异，几乎全发生在组（学校）内，因此[平均家庭作业]（$\bar{H}$）在各校之间就没有差异，带有[平均家庭作业]（$\bar{H}$）的跨层级交互作用 $\bar{H}H$ 与 $\bar{H}S$ 几乎等于原来的第一层的 H 变数，造成 S 与 $\bar{H}S$ 的回归系数会非常不稳定。不仅在原始数据时如此，在组平减数据中亦然。此外，S 变量的变异多存在于学校之间，因此 $\bar{S}H$ 与 H 会有很大的不同，而会接近 $\bar{S}$。

从相关系数的分析可以得到两个结论：第一，即使在固定系数模型，跨层级交互作用的使用会有很大的问题。在原始数据型态的模型中，甚至于组平减的模型，估计结果会十分不稳定。整体来说，组平减程序可以减缓多元共线性的威胁。第二层解释变量与第一层解释变量及跨层级交互作用变量两者之间的相关接近 0，表示我们只要担心跨层级交互作用与相对应的第一层解释变量的高相关问题就好了。

批 注

1. 该网络社群为 multilevel@ mailbase. ac. uk,想参与社群讨论者,仅需寄出一封主旨为:join multilevel your name 到前述地址即可,参与者可以张贴意见给所有会员。利用因特网者可由下列网址获得协助:http://www.ioe.ac.uk/multilevel 或 http://www.medent.umontreal.ca/ 与 http://www.edfac.unimelb.edu.au/multilevel/ 加以连结。
2. 假设回归方程式右项为 $\alpha+\beta x$,在 x 变量上增加一个常数 c,截距会变成 $\alpha-\beta c$,因为 $\alpha+\beta x=(\alpha-\beta c)+\beta(x+c)$。
3. 依循 Snijders 与 Bosker(1994)的说法,我们将使用"模式变异"(modeled variance)一词以避免混淆。
4. 这是一个讨论混淆原因的好机会,我们在第 3.9 节已经有所提及。在 RANCOVA 模型中,有一个基本的参数 τ^2,亦即随机截距 $\underline{\alpha}_j$ 的方差。此一方差反应了截距在不断独立重复实验后的变异情形,我们假设所有的截距都有相同的方差。在 ANCOVA 中,模型中则无截距方差,但是我们仍可以计算出 m 个固定截距的方差,用以反应截距变动的情形。最后,我们可以计算 m 个 ANCOVA 的估计数 $\hat{\alpha}_j$ 的方差,这些方差的概念都是有关联的,但是却不相同,说明了模型特性的差异,主要是看参数估计量的差别。
5. 从更一般性的情况下可以得到更多结果,但是看起来会非常复杂,或是需要用矩阵来说明。

附录　NELS88 数据库编码表

性别

Label	Code	次数	%
Male 男	1	10 564	0.49
Female 女	2	11 016	0.51

种族别

Label	Code	次数	%
Asian or Pacific Islander	1	1 277	0.06
Hispanic, regardless of race	2	2 633	0.12
Black, not of Hispanic origin	3	2 480	0.11
White, not of Hispanic origin	4	14 933	0.69
American Indian or Alaskan Native	5	257	0.01

每周做数学作业时间

Label	Code	次数	%
None	0	1 779	0.08
less than 1 hour	1	8 949	0.41
1 hour	2	4 942	0.23
2 hours	3	2 285	0.11
3 hours	4	1 653	0.08
4-6 hours	5	1 563	0.07
7-9 hours	6	262	0.01
10 or more	7	147	0.01

学校类型

Label	Code	次数	%
Public school	1	16 952	0.79
Catholic school	2	2 327	0.11
private, other religious affiliation	3	944	0.04
private, no religious affiliation	4	1 357	0.06

社经地位

Mean	-0.04
Variance	0.63

父母最高教育水平

Label	Code	次数	%
Did not finish H. S.	1	2 116	0.1

H. S. grad or GED	2	4 099	0.19
GT H. S. & LT 4yr degreee	3	8 627	0.4
College graduate	4	3 341	0.15
M. A. or equivalent	5	2 086	0.1
Ph. D. , M. D. , other	6	1 311	0.06

数学答对题数

Mean	51.01
Variance	103.72

班级环境结构性

Label	Code	次数	%
Not at all accurate	1	213	0.01
	2	439	0.02
	3	2 360	0.11
	4	10 588	0.49
very much accurate	5	7 980	0.37

学校规模

Label	Code	次数	%
1-199 students	1	1 045	0.05
200-399	2	4 331	0.2
400-599	3	5 404	0.25
600-799	4	4 666	0.22
800-999	5	2 911	0.13
1000-1199	6	1 584	0.07
1200 +	7	1 639	0.08

都会与城乡

LABEL	Code	次数	%
Urban	1	6 500	0.3
Suburban	2	8 998	0.42
Rural	3	6 082	0.28

学校所在地理区域

Label	Code	次数	%
NORTHEAST	1	4 246	0.2
NORTH CENTRAL	2	5 659	0.26
SOUTH	3	7 470	0.35
WEST	4	4 205	0.19

学校中的少数民族比例

Label	Code	次数	%
None	0	2 760	0. 13
1% ~5%	1	4 905	0. 23
6% ~10%	2	2 478	0. 11
11% ~20%	3	2 928	0. 14
21% ~40%	4	3 173	0. 15
41% ~60%	5	1 879	0. 09
61% ~90%	6	1 943	0. 09
91% ~100%	7	1 514	0. 07

生师比

Label	Code	次数	%
10 and below	10	1 451	0. 07
	11	780	0. 04
	12	599	0. 03
	13	1 514	0. 07
	14	1 665	0. 08
	15	1 895	0. 09
	16	2 002	0. 09
	17	1 486	0. 07
	18	1 924	0. 09
	19	1 426	0. 07
	20	1 043	0. 05
	21	1 161	0. 05
	22	800	0. 04
	23	1 009	0. 05
	24	522	0. 02
	25	523	0. 02
	26	455	0. 02
	27	241	0. 01
	28	294	0. 01
	29	239	0. 01
	30	524	0. 02

术语英汉对照表

（根据拼音排序）

Cronbach 模型 Cronbach model
Gauss-Newton 法 Gauss-Newton method
Gibbs 抽样法 Gibbs sampler
Robinson 效果 Robinson effect
“斜率结果”分析‘slope-as-outcome’ analysis
贝叶斯 Bayes
被解释变异 explained variance
变动截距模型 varing intercept model
变动系数模型 varying coefficient model
变化系数 changing coefficient
不稳定性 instability
参数化 parameterization
成长曲线分析 growth curve analysis
等值线性模型 equivalent linear model
地理信息系统 geographical information systems
调节变数 moderator variable
迭代 iteration
独特变异 uniqueness
多层模型 multilevel model
多元共线性 multicollinearity
反应变量 response variable
方差成分分析 variance component analysis
微观层次 micro level
固定变量 fixed variable
固定系数 fixed coefficient
固定效果 fixed effect
混合模型分析 mixed model analysis
混和线性模型 mixed linear model
基准 baseline
计分法 method of scoring
技术 techniques
简效性 parsimony
阶层 hierarchy
阶层数据结构 hierarchical data structure
结果 outcome
解释 description
解释变量 explanatory variables
解释变异量 explained variance
借力 borrow strength
经验贝叶斯 empirical Bayes
精确性 precision
聚合测量 aggregated measurement
聚合分析 aggregate analysis
聚合回归 aggregate regression
聚合脉络变量 aggregated context variables
跨层级交互作用 cross-level interaction effect
离差分数 deviation score
离差分数形式 deviation form
理解 understanding
理论 theory
两阶段分析 two-ste Panalysis
两权消长 trade-off
脉络 contexts
脉络变数 contextual variable
脉络分析 contextual analysis
脉络模型 contextual models
蒙地卡罗法 Monte Carlo
模式变异 modeled variance
模型 models
偏差度 deviance
频率混合模型 frequentist mixture models
起始值 starting value
嵌套 nested
扰动项 disturbance
设计变数 design variables
生态谬误 ecological fallacy
事后平均数 posterior means
收敛 convergence
双生子研究 twin study
算法 algorithms

随机变数 random variable
随机系数 random coefficient
随机系数模型 random coefficient model
随机效果 random effect
损失函数 loss function
缩动 skrinkage
缩动估计值 shrinkage estimator
同质性 homogeneity
统计技术 statistical technique
统计原理 statistical principle
完全信息最大似然法 full information maximum likelihood; FIML
限制性或残差最大似然法 restricted or residual maximum likelihood; REML
相关比 correlation ratio
协变量 covariate
协方差分析 analysis of covariance; ANCOVA
信度 reliability
性向与处理交互作用 aptitude-treatment interaction; ATI
虚无模型 null model
学校效能 school effectiveness
一般或加权最小平方法 generalized or weighted least squares; GLS
预测 prediction
元分析 meta-abalysis
整体回归 total or pooled regression
整体特征 global characteristics
整体效果 overall effect
追踪数据 panel data
自我回归的路径模型 autoregressive path model
自我相关 autocorrelation
宏观层次 macro level
总组间变异 total between variance
组别 groups
组内相关 intra-class correlation
最大似然法 maximum likelihood
最小平方回归分析 ordinary least squares; OLS

参考文献

Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interaction.* Newbury Park, CA: Sage.

Airy, G. B. (1861). *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations.* Macmillan, London.

Aitkin, M. A., & Longford, N. T. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society A*, 149, 1-43.

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, 6(3), 267-85.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and Social Psychology*, 51, 1173-82.

Bassiri, D. (1988). *Large and small sample properties of maximum likelihood estimates for the hierarchical linear model.* Ph. D. thesis, Department of Counseling, Educational Psychology and Special Education, Michigan State University.

Belsley, D. A. (1991). *Conditioning Diagnostic.* Wiley, New York.

Beran, R., & Hall, P. (1992). Estimating coefficient distributions in random coefficient regressions. *Annals of Statistics*, 20, 1970-84.

Boyd, L. H., & Iversen, G. R. (1979). *Contextual Analysis: Concepts and Statistical Techniques.* Wadsworth, Belmont, CA.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage Publications, Newbury Park, CA.

Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. T. (1988). *An introduction to HLM: Computer Program and User's Guide.* University of Chicago.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). *HLM: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs.* Scientific Software international, Chicago.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, 8, 158-233.

Burstein, L., Linn, R. L., & Capell, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics*, 3, 34-83.

Burstein, L., Kim, K. -S., & Delandshere, G. (1989). Multilevel investigation of systematically varying slopes: Issues, alternatives, and consequences. In R. D. Bock (ed.), *Multilevel Analysis of Educational Data.* Academic Press, New York.

Busing, F. M. T. A. (1993). *Distribution characteristics of variance estimates in two-level models. Preprint PRM* 93-04. Psychometrics and Research Methodology, Leiden. Netherlands.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2. North-Holland, Amsterdam.

Chow, G. C. (1984). Random and changing coefficient models. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2. North-Holland, Amsterdam.

Cochran, W. G. (1977). *Sampling Techniques.* Wiley, New York.

Cohen, M. P. (1995). *Sample sizes for survey data analyzed with hierarchical linear models.* National Center of Education Statistics, Washington, DC.

Coleman, J., Hoffer, T., & Kilgore, S. (1982). Cognitive outcomes in public and private schools. *Sociology of Education*, 55, 162-82.

Cressie, N. A. C. (1991). *Statistics for Spatial Data.* Wiley, New York.

Cronbach, L. J., & Webb, N. (1975). Between class and within class effects in a reported aptitude x treatment interaction: A reanalysis of a study by G. L. Anderson. *Journal of Educational Psychology*, 67, 717-24.

Davidian, M., & Gallant, A. R. (1 992). *NLMIX A Program for Maximum Likelihood Estimation of the Nonlinear Mixed Effects Model with a Smooth Random Effects Density.* Department of Statistics, North Carolina State University, Raleigh, NC.

de Leeuw, J. (1994). Statistics and the sciences. In I. Borg and P. Mohler (Eds.), *Trends and Perspectives in Empirical Social Research*, 139-48. Walter de Gruyter, Berlin.

de Leeuw, J., & Kreft, I. G. G. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11, 57-86.

de Leeuw, J., & Kreft, I. G. G. (1995). Questioning multilevel models. *Journal of Educational and Behavioral Statistics*, 20, 171-90.

Dielman, T. E. (1992). *Pooled Cross-Sectional and Time Series Data Analysis.* Marcel Dekker, New York.

Diggle, P. J., Liang, K. -Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Clarendon Press, Oxford.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics*, 20(2), 115-47.

Duncan, O. D., Curzort, R. P., & Duncan, R. P. (1966). Statistical Geography: Problems in Analyzing Areal Data. Free Press, Glencoe, IL.

Efron, B., & Morris, C. N. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 74, 311-19.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, 3, 1-21.

Engel, B. (1990). The analysis of the unbalanced linear models with variance components. *Statistica Neerlandica*, 44, 195-219.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399-433.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh and London.

Geisser, S. (1980). Growth curve analysis. In P. R. Krishnaiah (ed.), *Handbook of Statistics*, Volume 1. North-Holland, Amsterdam.

Goldstein, H. (1987). *Multilevel Models in Educational and Social Research.* Griffin, London.

Goldstein, H. (1995). *Multilevel Statistical Models.* Edward Arnold, London.

Hartley, H. O., & Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54, 93-108.

Harville, D. A. (1977). Maximum-likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320-40.

Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Models.* Chapman & Hall, London.

Hastie, T., & Tibshirani, R. (1993). Varying coefficient models (with discussion). *Journal of the Royal Statistical Society B*, 55, 757-96.

Hedeker, D. (1989). Random regression models with autocorrelated errors. Ph. D. thesis, University of Chicago.

Hedeker, D. (1989) Random regression inodels with autocorrelated crrors. Ph. IJ. thesis, University of Chicago.

Hedeker, D., & Gibbons, R. (1993a). *MIXOR. A Computer Program for Mixed Effects Ordinal Probit and Logistic Regression Analysis.* University of Illinois at Chicago.

Hedeker, D., & Gibbons, R. (1993b). *MIXREG. A Computer Program for Mixed Effects Regression Analysis with Autocorrelated Errors.* University of Illinois at Chicago.

Hedeker, D., & Gibbons, R. D. (1994). A random effects ordinal regression model for multilevel analysis. *Biometrics*, 50, 933-44.

Hemelrijk, J. (1966). Underlining random variables. *Statistica Neerlandica*, 20, 1-7.

Hemmerle, W. J., & Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed A. O. V. model using the W-transformation. *Technometrics*, 15, 819-31.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9, 226-52.

Hsiao, C. (1986). *Analysis of panel Data.* Cambridge University Press, Cambridge.

Jennrich, R., & Schluchter, M. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, 42, 805-20.

Johnson, L. W. (1977). Stochastic parameter regression: an annotated bibliography. *International Statistical Review*, 45, 257-72.

Johnson, L. W. (1980). Stochastic parameter regression: an additional annotated bibliography. *International Statistical Review*, 48, 95-102.

Kim, K. -S. (1990). *Multilevel data analysis: A comparison of analytical alternatives.* Ph. D. thesis, University of California, Los Angeles.

Kreft, I. G. G. (1994). Multilevel models for hierarchically nested data: Potential applications in substance abuse prevention research. In L. M. Collins and L. A. Seitz (Eds.), *Advances in Data Analysis for Prevention intervention Research.* Research Monograph 108. National institute on Drug Abuse, Washington, DC.

Kreft, I. G. G., de Leeuw, J., & Kim, K. -S. (1990). *Comparing four different statistical packages for hi-*

erarchical linear regression: Genmod, HLM, ML2, VARCL. Preprint 50. UCLA Statistics, Los Angeles, CA.

Kreft, I. G. G., de Leeuw, J., and van der Leeden, R. (1994) Review of five multilevel analysis programs: BMDP-SV, GENMOD, HLM, ML3, VARCL. *American Statistician*, 48, 324-35.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30, 1-22.

Lazarsfeld, P. F., & Menzel, H. (1969). On the relation between individual and collective properties. In A. Etzioni (Ed.), *A Sociological Reader on Complex Organizations*, pp. 499516. Holt, Rinehart k Winston, New York.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society B*, 34, 1-441.

Lindsey, J. K. (1993). *Models for Repeated Measurements.* Clarendon Press, Oxford.

Longford, N. (1993). *Random Coefficient Models.* Oxford University Press, Oxford.

Longford, N. T. (1990). *VARCL: Software for Variance Component Analysis of Data with Nested Random Effects* (Maximum Likelihood). Educational Testing Service, Princeton. NJ.

McMillan, N. J., & Berliner, M. J. (1994). *A spatially correlated hierarchical random effect model for Ohio corn yield. Technical Report* 10. National institute for Statistical Sciences, Research Triangle Park, NC.

Mok, M. (1995). *Sample size requirements for a 2-level designs in educational research.* Macquaric University, Sydney, Australia.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and application. *Journal of the American Statistical Association*, 78, 47-65.

National Research Council (1992). Combining Information, Statistical Issues and Opportunities for Research. National Academy Press, Washington, DC.

Pothoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

Prosser, R., Rasbash, J., and Goldstein, H. (1991). Data analysis with ML3. Technical report. Institute of Education, University of London.

Rao, C. R. (1965). The theory of least squares when parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, 52, 447-58.

Rasbash, J., and Woodhouse, G. (1995). *MLn Command Reference.* Institute of Education, University of London.

Rasbash, J., Prosser, R., and Goldstein, H. (1990). *ML3. Software for Three-Level Analysis. User's Guide.* Institute of Education, University of London.

Rasbash, J., Prosser, R., and Goldstein, H. (1991). *ML3. Software for Three-Level Analysis. User's Guide for V2.* Institute of Education, University of London.

Raudenbush, S. W., and Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *Sociological Review*,

15,351-7.

Rosenberg, B. (1973). A survey of stochastic parameter regression. *Annals of Economic and Social Measurement*, 2, 381-97.

Rubin, H. (1950). Note on random coefficients. In T. C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*. Wiley, New York.

Samuels, M. L., Casella, G., and McCabe, G. P. (1991). Interpreting blocks and random factors (with discussion). *Journal of the American Statistical Association*, 86, 798-821.

Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement*, 16, 209-22.

Scheffé, H. (1956). Alternative models for the analysis of variance. *Annals of Mathematical Statistics*, 27, 251-71.

Searle, S. R. (1979). Notes on variance components estimation. A detailed account of maximum likelihood and kindred methodology. Technical report BU-673-M, Biometrics Unit, Cornell University, Ithaca, NY.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. Wiley, New York.

Smith, A. F. M. (1973). A general Bayesian linear model. *Journal of the Royal Statistical Society B*, 35, 67-75.

Snijders, T. A. B., and Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods and Research*, 22, 342-63.

Speed, T. P. (1987). What is an analysis of variance? (with discussion). *Annals of Statistics*, 15, 885-941.

Spjφtvoll, E. (1977). Random coefficients regression models. A review. *Mathematische Operationsforschung und Statistik*, 8, 69-93.

Strenio, J. L. F., Weisberg, H. I., and Bryk, A. S. (1983). Empirical Bayes estimation of individual growth curve parameters and their relationshi Pto covariates. *Biometrics*, 39, 71-86.

Swamy, P. A. V. B. (1971). *Statistical Inference in a Random Coefficient Mode*l. Springer-Verlag, New York.

Tate, R. (1985). Methodological observations on applied behavioral science. *Journal of Applied Behavioral Science*, 21(2), 221-34.

Thomson, R. (1980). Maximum likelihood estimation of variance components. *Mathematiche Operationsforschung und Statistik, Serie Statistik*, 11, 545-61.

Timm, N. H. (1980). Multivariate analysis of variance of repeated measurements. In P. R. Krishnaiah (ed.), *Handbook of Statistics, Vol.* 1, North-Holland, Amsterdam.

van der Leeden, R., and Busing, F. M. T. A. (1994). First iteration versus IGLS/RIGLS estimates in two-level models: a Monte Carlo study with ML3. Preprint PRM 94-03. Psychometrics and Research Methodology, Leiden, Netherlands.

Velicer, W. F. (1972). The moderator variable viewed as heterogeneous regression. *Journal of Applied Psychology*, 56, 266-9.

Wald, A. (1947). A note on regressions analysis. *Annals of Mathematical Statistics*, 18, 586-9.

Wansbeek. T. J. (1980). *Quantitative effects in panel data modelling.* Ph. D. thesis, University of Leiden.

Weisberg, S. (1985). *Applied Linear Regression* (2nd edition). Wiley, New York.

Wilk, M. B. and Kempthorne, O. (1955). Fixed, mixed, and random models. *Journal of the American Statistical Association*, 50, 1144-67.

Woodhouse, G. (1995). *A Guide to MLn for New Users.* Institute of Education, University of London.